OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: May 3, 2003, 23:04:17 ; Search time 5082.23 Seconds
                                         {without alignments}
                                         17041.707 Million cell updates/sec

Title: US-10-027-000-1
Perfect score: 2976
Sequence: 1 ttatagtcgctttgttaat.........aaaaaaaaaaaaaaaaa 2976

Scoring table: IDENTITY_NUC
                 Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 1455140287B residues

Total number of hits satisfying chosen parameters: 4109280

Post-processing: Minimum Match 0%
                 Maximum Match 100%
Minimum DB seq length: 0
Maximum DB seq length: 200000000   Listing first 45 summaries

Database :   GenEmbl:*
        1: gb_ba:*
        2: gb_htg:*
        3: gb_in:*
        4: gb_om:*
        5: gb_ov:*
        6: gb_pat:*
        7: gb_ph:*
        8: gb_pl:*
        9: gb_pr:*
       10: gb_ro:*
       11: gb_sts:*
       12: gb_sy:*
       13: gb_un:*
       14: gb_vi:*
       15: em_ba:*
       16: em_fun:*
       17: em_hum:*
       18: em_in:*
       19: em_mu:*
       20: em_om:*
       21: em_or:*
       22: em_ov:*
       23: em_pat:*
       24: em_ph:*
       25: em_pl:*
       26: em_ro:*
       27: em_sts:*
       28: em_un:*
       29: em_vi:*
       30: em_htg_hum:*
       31: em_htg_inv:*
       32: em_htg_other:*
       33: em_htg_mus:*
       34: em_htg_pln:*
       35: em_htg_rod:*
       36: em_htg_mam:*
       37: em_htg_vrt:*
       38: em_sy:*
       39: em_htg_hum:*
       40: em_htgo_mus:*
       41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	%	DB ID	Description
1	427	14.3	4193	8	KFBGLUC	X05918 Kluyveromyc
2	406.2	13.6	38269	1	AF486581	AF486581 Uncultur
3	393.6	13.2	39305	8	SPU33009	U33009 Schizosacch
4	393.6	13.2	41799	8	SPBC1683	AL355920 Schizosac
5	393.6	13.2	85837	8	SPU33010	U33010 Schizosacc
6	392.6	13.2	2793	8	AF329731	AF329731 Volvariel
7	350.4	11.8	3578	1	AF005277	AF005277 Cellulomo
8	343.4	11.5	3710	1	ATUCBG1	M59852 A.tumefacie
9 c	305.2	10.3	2636	8	REU80928_2	AJ293760 Agaricus
10	270	9.1	110000	6	AX489445	AX489445 Streptomy
11 c	205.2	6.9	37445	1	SC8A6	AL031013 Streptomy
12	197.4	6.6	3241	6	AR173232	AR173232 Sequence
13	184.4	6.2	17665	1	AF521878	AF521878 Streptomy
14	184.4	6.2	2760	1	SEY14327	Y14327 Saccharopol
15 c	178.2	6.0	2760	1	AF079762	AF079762 Streptomy
16	175.2	5.9	2256	6	E11369	E11369 DNA encodin
17	161	5.4	2640	1	TNBGLB	Z77856 T.neapolita
18	157.2	5.3	3179	1	CEGDGB	D14068 C.gilvus ge
19	156.8	5.3	11763	1	AE005774	AE005774 Caulobact
20 c	149.4	5.1	14878	1	AE005849	AE005849 Xanthomon
21	149.4	5.0	10615	1	AE012240	AE012240 Xanthomon
22 c	143.6	4.8	15052	1	AF055579	AF055579 Streptomy
23	138.6	4.7	15052	6	AR159871	AR159871 Sequence
24 c	138.6	4.6	212243	7	AE001690	AE001690 Thermotog
25	137.4	4.6	3265	1	RAU92808	U92808 Ruminococcu
26 c	136.2	4.6	2166	8	AR205082	AR205082 Sequence
27	135.8	4.5	3158	1	RALBGLT	X15415 Ruminococcu
28	132.6	4.5	2075	1	ARGERMA	M11276 Arthrobacte
29 c	129.2	4.3	2620	1	CSBGL2	Z94045 C.stercorar
30	127.6	4.3	4537	1	AF090429	AF090429 Azospiril
31	127.6	4.3	2346	1	PAL276021	AJ276021 Prevotell
32	126.4	4.2	12140	1	AE011777	AE011777 Xanthomon
33 c	126	4.1	110000	2	LMFLCHR36_08	AB071166 Gluconace
34 c	123.2	4.1	3542	1	AB071166	AB071166 Continuation (9 of
35	118.2	4.0	3955	1	MIBBGLA	L06134 Microbispor
36	117.2	3.9	11557	1	AE004598	AE004598 Pseudomon
37	114.4	3.8	15852	1	AE006929	AE006929 Mycobacte
38	111.6	3.8	36300	1	MTC128	Z97050 Mycobacteri
39	111.6	3.8	13332	1	AE005788	AE005788 Caulobact
40	109	3.7	2370	1	PRU35425	U35425 Prevotella
41	106.2	3.6	2894	1	CTBGLB	X15644 Clostridium
42	104.4	3.5	12882	1	AE013715	AE013715 Yersinia
43 c	104.4	3.5	150754	9	AC023491	AC023491 Homo sapi
44	104.4	3.5	258050	1	AJ414153	AJ414153 Yersinia
45						

ALIGNMENTS

RESULT 1
KFBGLUC   KFBGLUC   4193 bp   DNA   linear   PLN 12-SEP-1993
LOCUS   KFBGLUC   4193 bp   DNA   linear   PLN 12-SEP-1993
DEFINITION   Kluyveromyces fragilis beta-glucosidase gene.
ACCESSION   X05918
VERSION   X05918.1 GI:2804
KEYWORDS   beta-glucosidase.
SOURCE   Kluyveromyces marxianus
  ORGANISM   Kluyveromyces marxianus
                 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
                 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1 (bases 1 to 4193)
  AUTHORS   Raynal,A., Gerbaud,C., Francingues,M.C. and Guerineau,M.
  TITLE   Sequence and transcription of the beta-glucosidase gene of
                 Kluyveromyces fragilis cloned in Saccharomyces cerevisiae

```
JOURNAL          Curr. Genet. 12 (3), 175-184 (1987)
  MEDLINE        88210533
  PUBMED         2835179
FEATURES                  Location/Qualifiers
     source            1..4193
                       /organism="Kluyveromyces marxianus"
                       /strain="ATCC 12424"
                       /db_xref="taxon:4911"
                       /clone_lib="PHCG3"
     repeat_region     268..276
                       /note="direct repeat 1"
     repeat_region     307..315
                       /note="direct repeat 1"
     promoter          316..320
                       /note="TATA-Box like sequence"
     promoter          342..355
                       /note="TATA-Box like sequence"
     repeat_region     400..408
                       /note="direct repeat 2"
     repeat_region     406..412
                       /note="direct repeat 2"
     promoter          416..424
                       /note="TATA-Box like sequence"
     misc_feature      459
                       /note="direct repeat 2"
     misc_feature      482..483
                       /note="pot. transcription start site"
     CDS               612..3149
                       /note="beta-glucosidase (AA 1 - 845)"
                       /codon_start=1
                       /protein_id="CAA29353.1"
                       /db_xref="GI:2805"
                       /db_xref="SWISS-PROT:P07337"
                       /translation="MSKFDVEQLLSELNQDEKISLLSAVDFWHTKKIERLGIPAVRS
                       DGPNGIRGTKFPDGVPSGCFPNGTGLASTFDRDLLETAGKLMAKESIAKNAAVILGPT
                       TNMQRGPLGGRGPESFSEDPYLAGMATSSVVKGMQGEGIAATVKHFVCNDLEDQRFSS
                       NSIVSERALREIYLEEPRLAVKHANPVCIMTAYNKVNGDDHCSQSKKLLIDILRDEWKW
                       QVLKMIKFVVDNLEKTGIVENGPESTSNNTKETSDLREIADSIVLLKNNNYLTSK
                       ERRQVHVIGPNAKAKTSSGGSASMNSYVVVSPYEGIVNKLGKEVDYTVGAYSHKSIG
                       GLAESSLIDAAKPADAENAGLIAKFYSNPVEERSEDEEPFHVTLKNRSNVHLFDFKHE
                       KVDPKNPFVPVTLTGQYVPQEDGYIFSLQVYGGSGLFYLNDELIIDQRHNQERGSFCF
                       GAGTKERTKKILTLKKGQYVNVRVEYGSGPTSGLVGEFGAGPQAGVIRAIDDDEEIRN
                       AAELAAKHDKAVLIIGLNGEWETEGIDRENMDLPKRTNELVRAVLKANPNTVIVNQSG
                       TPVEFPWLEEANALVQANYGGNELAGNALIAVDLYGDVPNGKLSLSWPFKLQDNPAFIN
                       VDVNTGDKFPAGSEVVQVFSALNSKVSRPVKELKGFEKVHLEPGEKKTVNIELELKD
                       AISYFNEELGKHHVEAGEYLVSVGTSSDILSVKEFKVEKDLYWKGL"
     misc_feature      3314..3322
                       /note="inverted repeat A"
     repeat_unit       3322
     misc_feature      3442..3445
                       /note="inverted repeat A"
     misc_feature      3472..3474
                       /note="yeast termination consensus sequence"
     misc_feature      3498..3501
                       /note="yeast termination consensus sequence"
     misc_feature      3541..3543
                       /note="yeast termination consensus sequence"
     repeat_unit       3642..3652
                       /note="yeast termination consensus sequence"
     repeat_unit       3642..3652
                       /note="inverted repeat B"
     repeat_unit       3655..3666
                       /note="inverted repeat B"
     repeat_unit       3714..3722
                       /note="inverted repeat A'"
BASE COUNT      1338 a    675 c    941 g   1239 t
ORIGIN
```

```
QY   339  ATGGCTGGATAATTGANGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCGGAGAAGTCGAT 398
          ||| || | ||||||||| | ||||||||||||||  ||||  ||||||||||||||||
Db   612  ATGTCTAAATTTGATGTTGAACAGTTATTGAGTGAATTGAACCAGGATGAAAAGATTCC   671

Qy   399  CTGGCTGGGCTGGTATCGACTTCTGGCACACAAAGGTCTCTCCCCAAGCCATGGAGCCCTCT 458
          ||| |||||  ||    || ||| ||| | |||  | || | |||||||||||
Db   672  TTACTTTCTGCAGTTGATTTCTGGCATACTAAGAAGATTGAAAGACTTGGGATTCCACGG   731

Qy   459  CTCCGCTTTACAGATNGCCCCCAAACGGCGTAAGAGGCGACCAAGTTCTTCAATGGCGTCCCT 518
          ||| ||||||| | |||||| |  ||||||  |||||||| ||||||  || ||| |  ||||
Db   732  GTGAGGGGTTTCTGATGGTCCAAAATGGTATTAGAGGGACGAAGTTCTTTGATGGGGTTCCT   791

Qy   519  GCGGCCTGCTTCCCTTGCCGCACGTCGGTTCCACATTCAACAAAATTCTGCTCGAA   578
          || ||| |  ||||| || |||| | ||| |  |||||||||| |||| |||||
Db   792  TCAGGATGTTTCCCTAATCGTACCGGCGTTGGCAATCACTTTTGATCGTCGTGAG   851

Qy   579  GAGGCAGGTAAGACTGAATGGGCAAAGAGGCCATCGCTGAAGACTGCGCCAT   638
          ||||  |||| || |||  |||||  | ||| |  ||||||| ||||| |||
Db   852  ACAGCAGGTAAGTTGATGGCTAAGGAAATCGATTGCCAAGAATGCTGCTG   911

Qy   639  CCGACTATCAACATGCAAGGCTCCCCTTCTCGGTGACGTGGGTTCGACTGGATTGGTGAG   698
          ||||| |  ||| || |||| ||| | ||||||| ||||| |||||| || ||| | |
Db   912  CCAACCACAAACATGCAAGCGTGGTCGTTCCTTGGGGTGGTCGTTGTTTTGAAATCATT   971

Qy   699  GATCGGTTCCTGCGGCGGCTTGGGAACTGCGCGCTCTCATCGCGGCATTCAGAGCACTGGA  758
          ||  |  |||| |  |  || || |||| | |||| || |  || | |||||||||||||
Db   972  GATCAATATCTTGCGTCGGTAGGCGTACTTCTCGTGTGTTGTTAAGGTATGCAAGGCGAAGGT  1031

Qy   759  GTGCAGGCTAGGATCAAGCAACACTTTTGTGCAAATGATCAGGAGGACAGAGGGCCATGATGGTG  818
          || ||||||| || || |  | ||| | |||  | |||| |||||| || |||||| ||||
Db   1032 ATTGCTGCTACCGTTAAGCAATTTGTTGTAACGAACTTGGAAGAACCTCGTTTCTCG   1091

Qy   819  CAGAGCAGTCGTCACGGAGCACGGGGCTCTCCGTGAAAATCTACGCACTCCCGTTCCAGATTGCT   878
          || ||| | ||| | |||||||| |||||  ||  ||||  ||| |  ||||| |||||||
Db   1092 AACTCAATTGCTACCGTTAAGCATTTGTTTGTAACGAACTTGGAAGAACCTCGTTTCTCG   1151

Qy   879  GTGCGGAGACTCCCAGCGCCGGGGTGCGTTCATGACGGCGTACAACAATGGCGTGTGG   938
          ||||||| | |||||  |  |||||| ||| |  ||||||||| |  ||||| ||||
Db   1152 GTTAAACATGCCAAATCCTGTTTGTATAAATGACCTGTTCTTATTAACAAGTCAATGGCAATCAT   1211

Qy   939  TGCAGCGAGAACCCTAAAATATCTTGATGGGATGCTTCGAAAAGGAATGGGGTTGGGATGGC   998
          |||||||  | |  ||  |  | |||||| |  ||||| ||||||| ||||||
Db   1212 TGCTCCCACAATCCAAGAAGTTAATGATCGACATTTGAGAGACGAGTGGAAAATGGGAACGGT   1271

Qy   999  CTAATCATGAGCGAGCTAGTTACGGCACATACACTACCACAAGCCGTTGTGTTGGGAGGGCCTG  1058
          ||||| || |   || | |||| |||| ||| || ||| | ||||| ||| || |||| |
Db   1272 ATGTTAAATGTCCGACTGGTTCGGTACAATATACGACTGGTCACANTATCAAGAATGGGGTG  1331

Qy   1059 GACCTGGAGATGCCCGACCTTCCAGGCGTTCCGGAGGAGAAACACTCAAGTTCAACGT----  1114
          ||| ||||| | |  ||||||| |  || ||  || ||||  | | ||||||||
Db   1332 GATATCGAGTTCCCTGACCAACACAGATCGAAGAACACGTGCTTTAGTGTCTCACTCTCTC  1391

Qy   1115 --CTCCAACGGAAAAGCCCTTTATTCCACGTCATTGACCAGAGGGGCTAGGGAGAAAGTTGTTAAA  1172
          ||  || |||  |||  ||| ||| || | |||  || |||||| || ||| |  ||||  ||
Db   1392 AACTCCAGAGAACAAAATCACTACTGAAGATGTTGATGATCCGTTAGACAAGTGCTAAAA  1451

Qy   1173 TTCGTCAAGAAGTGTTGCTCGCCTCCGGAGTGAC----------GGAGAACGGCGCCCCGAG  1220
          || |  ||||||||| | |||| ||| ||||| |         ||| ||||||||||||||
Db   1452 ATGATTAAGTTCGTTGTGTGACAATTTAGAGAAAACAGGTATTGTGGAGAAATGGGCCAGAA  1511

Qy   1221 ACGACTGTCAACAACACCCCCGAAACGGCAGCTGTCCTCCGGAAGGTTGGCAACGAGGGC  1280
          || ||||  || ||||| || |  | ||| || | || ||||  ||| |||| |  || |
Db   1512 TCTACTTCAAACAACAACACCAAGAAACCTCGGAACTGTTGAGAGAGATTGCTCGACTCT  1571

Qy   1281 ATCCTGCTGCTGAGAACGAGCAGACACACCCCGAAAGAGAGAGAGAGAGAGACGACGTG  1340
          ||| | |  |||| | ||||| |||||| | | ||||||||||||||||| ||| ||
Db   1572 ATTGTTTTTATTGAAGGACAAATAATTATTCCACCTCTAAAGAAAGACAAAAATCAT  1631

Qy   1341 ATTGTCGGCCCCAACGGCCAAGCAGCCCACATACCACGGCGGAGGCTCTGCCGCACTCAGG  1400
          || ||||||||| ||||| | || | ||  || | |||||||| || |||||||||||||
Db   1632 GTCATTGGCCCAAATGCTAAGAAAGACTAGTTCCGGTGGTGTTCAGGATGTTCCGGCACTC  1691

Qy   1401 GCCTACTACGCAGTCACTCCCTTTGACGGGCTCAGCGAAGCAGCGTCGAGACGACGCCCATCG  1460
          |||| ||| ||||  | ||| | | ||||| | || ||  || |  |||||||||||||||
Db   1692 TCCTACTATGTTGTTTCTCCGTATGAAGGTATCGTCAATAAGCTGGGCAAAGAGGTCGAT  1751
```

```
Qy   1461 TACACCGTCGGCGGGCCCTA--CACCACCGTCGTTCCTCCCATTCTAGGCGGAGCCAGGTGCCTCACGC 1518
          |||||||  ||||||||                ||||||
Db   1752 TACACCGTAGGGGCGGGCCTATTCACACAAAATCGATTGGAGGTTTGGCAGAGAGTAGTTGATC 1811

Qy   1519 CCGAGCGGGGCGCTCCGGCGCATGCGCCTGGAGGGTCGGATGC-GCCCCCTGGTTACCCCTAA 1576
          ||| |||| |||   ||                        ||    ||||
Db   1812 GATGCTGCAAAAACCAGCAGAGTGCTGGAAATGCTTGGATTAATTGCCAAGTTTTACTCAAT 1871

Qy   1577 CCGCCAGCACA------------TTGACGAGCTCTTCTTCACCAAGACG 1613
          ||||||||||              |||  ||||  |||| |
Db   1872 CCAGTAGGAAGAGAGAGATCTGAAGATGAAGAAACCATTCCACGTTACCAAAGTCAATAGATCC 1931

Qy   1614 GACATGCAACCTGGTTGGACTACTACCACCCCACACACCCCAGGCGGCGCCAGAGACG 1664
          ||||  |   |||   ||        ||    |           |||||
Db   1932 AAGTTCACTTATTTGATTTCAAACATGAGAAAAGTGGAATCCCAAAGAAACCCTTACTTTTTT 1991

Qy   1665 GCCGACATGGAGGCGCACGTACACCCCGACGAGGACTGCACCTTACGAGCTCGGCCTCGCTC 1724
          ||                |                      |
Db   1992 GTAACCTTAACCGGACACGTACGGGCCTCAAAGAGAATGTTGATTTATATCTTCAGTCTTCAA 2051

Qy   1725 GTCTGCGGGCCACGGCGAAAGGCGTACGTAGACGACCAGCTCGTCGTCGACAACGGCCACCAAG 1784
          |||
Db   2052 GTTTATGGTTCTCGGTTTGTTGTTCTACTTAAACGATGAGTTGATTAAGACCAAAAGCACAAC 2111

Qy   1785 CAGGTCCCCGGCGGATGCCGTTCTTCTCCGGGCTCCTGGGTCCGCACCCACCCACC 1844
          |||||||||||||||||||||      |     ||       |||  ||||    |
Db   2112 CAAGAAAGGGGTAGTTGTTCTGCTTTGGAGCTGGTACCAAAGAAAGAACCAAAAGTTGACT 2171

Qy   1845 CTCGTCAAGGGCAACACGTACACAGTTCAAGATCAGTTGGGCTCCACCACCTACACC 1904
          |||||          |||||      |||||||   |||||||  |||| ||   |
Db   2172 TTGAAGAAGGGCCAAGTTTTATAATGTAAGGAGTACGGTTGCTTGGGCCCAAACTTCAGT 2231

Qy   1905 CTCCAAGGGCGGACACCATCGTCCCCGGCCACGGCTTCCCTCCGGGCGGCGCTCGCAGGTC 1964
          |||||||||||||     ||
Db   2232 ------ATTCGGTCAGGTGGATTCCAAGCTGGTGTCATTCATTAAGGCG 2282

Qy   1965 ATTGACGACCAGGCCGGCAAATCGAAAAGTCCGTCGCCCTCGGCCAAGAGCACGACCAGGTC 2024
          |||||  ||||  |||||||      |||||||      |||   |  ||||  |||
Db   2283 ATCGATGATGACGAGGATTAGAAAACGCAGAATTAGCAGCTAAGCATGAAGAAGGCT 2342

Qy   2025 ATCACTCTGCCGCGGGCCTTAACGCACCCGACTCGGAGACCAGGACCCGCGAGCATG 2084
          ||||       ||||||||||    |||||||     ||      ||  ||||
Db   2343 GTGTTGATAATTGGATTAAAATGGTGAAAGAAACCGAAGGTTATGACAGAGAAAACATG 2402

Qy   2085 AAGCTCCCCGCGGGCGCTGGGACCCAGCTCATTGCCAGCGAACCCAAACACC 2144
          |||    |                ||||  |||||||||      ||||||||
Db   2403 GATTTGCCAAAAAGAACAAAATGAATTAGTTGCTGTCTGTTTTGAAAGCAAAATCCAAAAATACT 2462

Qy   2145 GTCGTCGTCATGCAGAGCGGGCAACCCCGAGGAGAGATCGCCTGGCTGCGACAGCGCACGC 2204
          |||||||||||||            ||||||      ||||||||||  ||||
Db   2463 GTTATCGTTAACCAATCAGGTAACCCCAGTTGAGTTCCCTTGGTTAGAAGGCAAAATGCG 2522

Qy   2205 GTCATCCAGGCCTGGTACGGCGCCGGCCAACGAGAGGGCAAACTCCATTGCCGACTCGTCTTT 2264
          |||||  |||||||||||||||||| ||||||   |||
Db   2523 CTAGTTCAAGCTTGGTGGTACGGTGGTAATGAATTGGGGTAATGTATCGCAGATGTCTTGTAC 2582

Qy   2265 GGCGACTACACACCCTCGGGCAAAGCGTGTCCTCCTCAGCTTCCCAAGCGCCTCGAGGACAAC 2324
          |||||||||||||| |||||||| |    |||||||||      |||||  |||
Db   2583 GGTGACGTGGTTCCAAATGGTAAGTTATCGCTCTTGGCCATTAAGTTGGCAAGATAAT 2642

Qy   2325 CCCGGCGGTTCTCAACTTCCCGCACCGAGGCGCGTCGGGACGGTCTAC 2384
          |||||  |||| ||||      ||||| |||||||      ||||||
Db   2643 CCAGCCTTTTTAAACTTCAAGACGAGTTGTTGTTCACGGGATATCTTT 2702

Qy   2385 GGTGGGTACAGGTACGTTGCCGGACAAGGACGTCAATTATTCCCCTTTTGGCACGGC 2444
          ||    |||||||||          |||||  || ||    ||    ||||
Db   2703 GTTGGTTATAGGTACTACGAAAAAGCTTCAAGAAAGGTAGCCTTCCCTTCGGATATGGT 2762

Qy   2445 CTGTCCTACACACAGACTTTTGCCTTTTCAAATCTCTCCGGTGTCTCACAAGGACGGCAAGCTG 2504
          ||||          |||   |          ||||      |     ||||||  |||||
Db   2763 CTATCGTTATACAACACATTCGAACTAGAATATTTCTGACTTCAACAAGGTAACTGATAGATA 2822

Qy   2505 AGGGTGTCCCTCCGTGGAAGAACACACCGGGTCCGTGCCCGGCCCGGCGCCACAGGTGGCCCAGCTC 2564
                |||   ||||||||||||   ||    |        |||     ||  ||||  |||
Db   2823 GATATTCAGTTGATGTGAAGAAATACTGGTGATAAAATTGCTGCCTCCGAGGTGGTTGCAA 2882

Qy   2565 TACGTCAAGGCCCCTCCAAGGCCGCCAAGGCGGCAAGATTAACCGCCCCGTCAAGGACGTTC 2624
          ||||| |||||  |   |||||||  ||||||| |||||||   ||  ||||||||||||
Db   2883 GTCTACTTCAGGTACTCTCTAAACTCTAAGGTCTCGAGACGCCCGGTTAAGAGTTGAAGGGATTC 2942

Qy   2625 GCAAAGGTCGACACTGCAGCCCGCGCGAGACGAGCAGGCGGGTGACAATCGAGAGGAGAAG 2684
          ||| |||||||      |||||||  |||||| ||||||  ||  ||||| |||||||||
Db   2943 GAAAAAGTTCATTGGAACCAGGTGAGAAGACAGTTAATATTGAAGAATTGAA 3002

Qy   2685 TACGTCCCGTGCCGTATTTTGATGAGGACCCGGACCGGGATCAGTGGTGTGTCGAAAAGGGC 2744
          ||||||  ||||  ||||||||||||   |  |||   ||||  ||||||||||||  |||||
Db   3003 GATGCAATTTCCTACTTTAACGAAGACAGTCGGTAAAATGGCACGTTGAAGCAGGTGAATAC 3062

Qy   2745 GAGGTTATCGT 2755
          ||||||||
Db   3063 TTGGTTTCAGT 3073
```

RESULT 2

```
LOCUS       AF486581               38269 bp    DNA     linear   BCT 21-MAR-2002
DEFINITION  Uncultured bacterium aminoglycoside 3'-N-acetyltransferase gene,
            complete cds.
ACCESSION   AF486581
VERSION     AF486581.1  GI:19569565
KEYWORDS    .
SOURCE      uncultured bacterium.
  ORGANISM  uncultured bacterium
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 38269)
  AUTHORS   Cappelano,C.M., Courtois,S., Ball,M., Francou,F.-X., Normand,P.,
            Helynck,G., Guerineau,M., Martinez,A., Hopke,J., Kolvek,S.,
            August,P.R., Osburne,M.S., Jeannin,P., Simonet,P. and
            Pernodet,J.-L.
  TITLE     Recombinant Environmental Libraries Provide Access to Microbial
            Diversity for Natural Products Drug Discovery
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 38269)
  AUTHORS   August,P.R. and Martinez,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Molecular Biodiversity, Aventis Cambridge
            Genomics Center, 26 Landsdowne Street, Cambridge, MA 02139, USA
FEATURES             Location/Qualifiers
     source          1..38269
                     /organism="uncultured bacterium"
                     /db_xref="taxon:77133"
                     /note="isolated from soil"
     CDS             complement(30829..31617)
                     /note="confers resistance to kanamycin in Escherichia
                     coli; similar to known cognate proteins"
                     /codon_start=1
                     /transl_table=11
                     /product="aminoglycoside 3'-N-acetyltransferase"
                     /protein_id="AAL92107.1"
                     /db_xref="GI:19569566"
                     /translation="MTSRVATRSSLADDLSAIGLADDGDAVLVHHALRQVGKIVGGPDA
                     IIDALRDVIGPAGTILGYCDWQLEDEDLRDDPSMRPHIAAFDPERSRSTRDNGYWPEAL
                     RTTPGALRSGSSPGASMAALGGEAEWFTADHALDYGYGPQSPLGKLVEAECKVLMGAP
                     LDTMTLLHHAEHLADFPNKRILRYEAPILVDGETVWRWFEEFPDYSEPPDGLPEDYFAT
                     IVEAFLATWGRGKRGEVGEASSVLVPAAAWVAFGVDWLERNGKTL"
BASE COUNT     6318 a   12117 c   12998 g   6835 t      1 others
ORIGIN
```

```
Query Match           13.6%;  Score 406.2;  DB 1;  Length 38269;
Best Local Similarity 50.8%;  Pred. No. 4.8e-44;
Matches 1213;  Conservative    0;  Mismatches 1098;  Indels 78;  Gaps 7;

Qy    355 TTGAGGCCATCTTGAAGAAGGTCACCCTGGCCCGAGAAGGTCGATCTGCTGCTGGTATCG 414
          |||||        ||||||||  ||  |    ||   |||||||   ||| ||||
Db  32244 TCGAGGAACTGGTCGACCGGATCGACGCTCGAGGAGCAGGTGAGCCTGTTATCGGGGCGAGG 32303
```

```
Qy    415 ACTTCTGCACAAAAGGCTCTCCCAAGCAATGGAGTNCCCTCTCTCCGGCTTTACAGATG 474
          |||||||
Db  32304 ATTTCTGGTTCGCTGCCCGCGATCAGATCGGTCGGCTGCGCGAAGTCCGGGTCACCGACG 32363

Qy    475 GCCCCAACGGCGGCTTAAGAGGGACCAAGTTCTTCAATGGC---GTCCCTGCGGGCCTGCTTCC 531
          ||||||                                                      |||||
Db  32364 GGCCGAACGGGGCGCGTGGCGCGGCTCGCGTCGATCGGGTCAAGTCAGGTCGGTTCC 32423

Qy    532 CTTGCCGGCACGTGCCTCGGTTCCACATTCAACCAAAAACTCTGCTCAGAAGAGGCAGGTAAGA 591
          |||||||
Db  32424 CGGTGGGGATCCGCGCTCGGGCCCAGCTGGGAACGTCCGGCTTGCTTACCGAGGTCGGCACGG 32483

Qy    652 TCCAACCCTCCCCTCTCCGGTGGACGTTGGCTTCGAGTTCGAGTTCGAGAAGATCCGTTCCTGG 711
          |||||||
Db  32544 TCCATCGTTCGGTTCAGCGAACCGGCGCAAATTTCGAGTGTTATTCCGGAGGACCCGAATCCTCA 32603

Qy    712 CGGGCTTGGGAGACTGGGGCTCTCATCCGCGGGCCATTCAGAGCGACTGCAGGCTACGA 771
          |||
Db  32604 CCGGCCAGCCTGGGCGTCGGCTTATATCGCAGGGGGCAGGGGTTGCAGGCCGACGA 32663

Qy    772 TCAAGCACTTTTTTGTGCAAATGATCAGAGGAGCACGGCGCATGGTGCAGAGCAGCCATCGTCA 831
          |||||||
Db  32664 TCAAGCACTTTGCCGGGCAAATGAAATCCGAGATCGAACCACCAGCTCCGAGATCG 32723

Qy    832 CGGAGCGGCGGCTCTCCGGTGAAATCTTACGCACTCCCGTTCCAGAATTGCTGTGCGAGACTCCC 891
          ||||||||
Db  32724 ACGAGCCGGGCGGGCGCTGCCTACCTCGTCGAGTGTGTTATTCCGGAGGACCCGAATCCCTCA 32783

Qy    892 AGCCGCGGGTGCCTTCATTGACCGGCGGTTACAAAAATGGCATCAAATGCGTGTGCGACGGGAGAACC 951
          ||| |||
Db  32784 GCACCTGGGCGGCATCATTGAGCTCCTTACAACAAGCTCAATTGCACCTTCACCACGAGAAACC 32843

Qy    952 CTAAATATCTTGATGGGGATGCTTCGAAAAGGAAATGGGGTTGGGATGGCCTAATCATCAGACCG 1011
          ||||||||
Db  32844 CCTGGCTGCTCGACACAGTTGCTCGGCACGGGAAATGCACTATGACGGCCTGGTGATGTCGG 32903

Qy   1012 ACTGGTACGGCGCACATACAGTACTACCACAGAGAGCCGTTGTGGCGCCTCGAGAATGC 1071
          |||||| |||
Db  32904 ACTGGTTCGGCTCGGCTCCGACTCCGACCCGCCCGGCCTCAATGGCCGCTGCAAATGC 32963

Qy   1072 CCGGACCTTCACGCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCT 1131
          ||| |||
Db  32964 CCGGGCCGGCCGCCGGACCGTGTGGCGGGAAGCTTGTGGCCGCCGGAGGTCA 33023

Qy   1132 TTATCCACGTCATTGACCAGAGGGCTAGGGAGAGTTCTTCAGTTCGTCAAGAAGTGTGCTG 1191
          |||
Db  33024 GCCGCGAAAGCGCTGCGGACCGGGTGCTCGACCTGCTTGAGGGGAACCGGGG 33083

Qy   1192 CCTCCGGAGTGACGGAGAAGAAGAGACGCTGTAGGGCCCCGAGACGACTGTCAAACAACCCCCGAAAACGGCCAC 1251
          ||||||
Db  33084 CGCTGGATCCGTCGCCGCCGGAAGGCGCGGGTGCTGTGCTGAAAAATGCGGGCGCGAACTGC 33140

Qy   1252 CTCTCCTCCGGAAGGTTGGCAACGGAGGGCATCGTGCTGACTGAAGAACGCAGCACAGGTTC 1311
          |||
Db  33141 CGCTGGATCCGTCGCCGCGCCGGGAAGGCGCGGGTGCTGTGCTGAAAAATGCGGGCGCGAACTGC 33200

Qy   1312 TGCGCCTTGAGCAAGAGAAGAAGAGACAGCTGTAGGGCCCCCAAGCGAAGCGGCCACAT 1371
          ||||||||
Db  33201 TGCCATTGAGCGGGGTCGGGCAACGCTAAATGGGGCGCAGA 33260

Qy   1372 ACCACGCGGGCGGAGGCTCTCGCCGACTCAGGGCCTACTACGCAGTCACTGCTCCCTTTGACGGGC 1431
          |||||||
Db  33261 TGCCAITGGGGCGCGGGTCGGGCAACGCTAAATGGGGCGCAGA 33320

Qy   1432 TCAGCAAGCAGCTGAGACGCGCCATCGTACACCGTCGGGGCCTTGGGCAAGAGAAGGCCCTTACACCACCGTTCCTC 1491
          |||
Db  33321 T-----------GGTCGCGGGCCTTGGGCGAAGAAACGCCCTCAGCTATGCGG 33359
```

```
Qy   1492 CCAITCTAGGCGAGCCAGTGCCTCACGCGGCCCTCACGCCGGGCCATGCGGCTGGAGGGTCT 1551
          ||| |||
Db  33360 CGGGTGCCGACCAACCACCGGCTTCGCGGGGCCTTGCGGCCTTGGCGTTCACGGCTGAGTGTT 33419

Qy   1552 TCAACGAGCCCCCTGGTACCCCTTAAACCGCCAGCACATTGACGGGGCCTGTCTTCTTCACCAAGA 1611
          |||
Db  33420 TCGCCAATGACAGACTCTCGCGGCCGCGGTCGGGCCAGCACCTGGGGGGCCTTCCACGGCTCACCAAGA 33479

Qy   1612 CGGACATGCACCTGGGTGGACTACTACCACCCCAAGGCCCCAAGGTGGTTACGCGACA 1671
          |||||| |||||||
Db  33480 CGTTCTGGATCGGGCGGACGACTCGGGTGGGGGCAAGGTCGGATCCAACTCACTGCATTTCTGCCAGGC 33539

Qy   1672 TGGAGGGCACGTACACCCGCGACGAGAGACTGCACCTTACGAGCGTGGGCCTTCGGCTGCGG 1731
          |||
Db  33540 TCACCGGCCGGTTCACCCCGCGAGGCGAACATCGGGTCGCATCTACAGCGCCG 33599

Qy   1732 GCACGGGCAAAGGCGTACGTAGAGACGACCAGTCGTGGTCGTGTCGACAAGCGAGGTCC 1791
          |||||||
Db  33600 GTTTTTCGCGGGTGCTGGGTCAGACGGCAAGGTGGGGCGGGATTGGGTCCCACCAAGGTCC 33659

Qy   1792 CCGGCGATGCCTTCTTCGGGGCCTCCGCACCCCGCGAGGAGGACGGCCGCCATCAATCTCGTCA 1851
          |||      |||
Db  33660 AAGGCCGCCACTTTCTTTGAGGAAAGGCTGCCGACGAGGTCGTCGGGGACGAATCGGAGTCCGAAG 33719

Qy   1852 AGGGCAACACACTACAGTTCAAGAGATGCAGGTCGGCTTCCCGCCACCCTACACCCTCAAGG 1911
          ||||||||
Db  33720 CCGGCCGCAGCTACGAAGTGGTGATGTCGGGTGATCGAGTTCGACGCGTTCGGGACGCTGGGGC 33779

Qy   1912 GGGACACCATCGTCTCCCGGCCACGGCCTCCCGCTCCGGGGGCTGCAAGGTCATTGACG 1971
          |||     |||
Db  33780 TCGCGGGCCTTCGCCCGGCCATCGGCGCTTGCCGGCCTTGGGCAC------------ 33820

Qy   1972 ACCAGGCCGCGAAAATCGAAAAGTNCCGTGCGCCCTCGCCCAAGGACACCAGGTCATCACCT 2031
          |||||| |||    |||||||
Db  33821 -------GCGGCGGATCGCCGAAGCGTTCGCACGGCGTCGACGGCGGCCGGGTGGTAT 33875

Qy   2032 GCGCGGGGCCTTAACGCCGACTGGGAGGCGCCGGAGGCGCGGCCGGTACGACCGAGCATGAAGCTCC 2091
          |||
Db  33876 TCGTCGGGCCGGCCAACGCGGGCGAGTGGGGAGACCGAGGCGAGCGACGCCTGATTCGGGGC 33935

Qy   2092 CCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCGCCGAGACCCAAACACCGTCGTCG 2151
          ||||||||
Db  33936 CCGGCCGGCAGGACGAGCTGATCGCGCGCGGTTGGGGTGGTGGAACCCGGTGGTGG 33995

Qy   2152 TCATGCCAGACGGGGCCACCCCCCGGAGGAGATGCCCGTGGGGTCGCGACCGCCCGGCATCC 2211
          |||||||
Db  33996 TGCTGCAGACCGGCGCGGGCGGCGGTGGAATCCGGTGTGGTCTCGGTGGACGGTCCCCTCG 34055

Qy   2212 AGGCCCTGGTACGGCGGCGCAACGAGGACGGGCGGCGAAACTCCATTGCCGACGACT 2271
          |||||||
Db  34056 AGGCCCTGGTATCGGCGAAAAATCTATCCCGGGCCTCAGACGGCGAAGGTGGCGCTACGAGGGG 34115

Qy   2272 ACAACCCCTCGGCGCAAGCGTGTCCCTCAGCTTCCCCCCAACCGCCTGCAGGAGCAACCC 2326
          |||||||
Db  34116 CCGAGCCGGCGGGCCTGCCGGCCAGCTTTCGGCGAGGTTCCCGGTCAAGGTGGCCGGACGACAACCA 34175

Qy   2327 -----CCGGTTCTCAACTTCCGCACCGAGGCGGGCGGCCGGGTTACGCGGGCGGAGGACG 2379
          |||||||
Db  34176 ACTCCGCAGGATNCGGCGAAATCTATCCCGGGCCTCGACGGCAAGGTGCGCTACGAGGAGGGG 34235

Qy   2380 TCTACGTCGGGGTACAGGTTACTACGAGTTTGCCGACAAGGAGGACGTCAATTTCCCCGTTTGGGCC 2439
          ||||||||
Db  34236 TGTTCATCGGCTATCGCCACACTACGATCGCCATGGCATCGGCGTTCCGGTCAAGGTGGCGGT 34295

Qy   2440 ACGGCCTGTCTACACCACTTTTGCCTTTTCCAAATCT-------------CTCCGTGTCTC 2487
          ||| |||      |||
Db  34296 TCGGACTTGGCTACACACCAGCTTTGACTGTCGGAGTTGAACTGCGAGCCTCGGGCTTCG 34355

Qy   2488 ACAAGGACGGCACAAGCTCAGGCGTGTCCCTCTCCGTGAAGAAGACACCGGTCGAGAAATACCGGCGGGGCGCGCCCGGG 2547
          ||| |||
Db  34356 AAGCGGATGGCAGTGTGAGGGTCTCGCGTCACGGTCGAATACCGGCGGGCGGGCGCGGCT 34415

Qy   2548 CACAGGTGGCCCAGCTCTACGTCAAGCCCTCCAAGCGGCCAAAGATTAACCGGCGCCCCGTCA 2607
```

```
Db  34416  CGGAGGTGGGTGCAGGTGCTACGTGTCTCGGCGCGGATGCG---ACGGTGTCGGCGGCGGCCA  34472
                |||  |||||||  |||    ||||||    ||||  |||||      ||||||||   |||
Qy  2608   AGGAGCTTCAAAGGCGTTCGCAAAAGGTCGAACTGCAGGCGCGGAGACGGAGAAGGCGGGTGACAA  2667
           ||||  |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  34473  AGGAGCTGAAGGCTTTCGAGAAGGCTCGCATCTGGCGCGGAGGCGCGCGGGGTGACGC  34532
           |||||||  ||||||||||||||||||||||||||||||||||||| |||||||
Qy  2668   TCGAGGAGCAGGAGAAGTACGTCGCTGCGTATTTGATGAGGAGCGGGGA  2716
           |||||||||||||||||||||||||||||||||||||||||||
Db  34533  TGGAGCTCGATGACCGGCGGCCTTTCGCTTACTACCGCACCGAAGCGCAGCA  34581
```

```
Query Match  13.2%;  Score 393.6;  DB 8;  Length 39305;
Best Local Similarity  49.0%;  Pred. No. 2.1e-42;
Matches 1199;  Conservative  0;  Mismatches 1229;  Indels  18;  Gaps  5;

Qy  339   ATGGCTGATATTGATGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAGGTCGAT  398
          ||||||| ||||||||||||||||||||| |||||||| |||| |||||||||||||
Db  8497  ATGATGGAACATGATGTTGAGATTTGATCAATCAATTGGATAAAGTCAGAAGGCTATG  8556

Qy  399   CTGCTGGCCTGGTATCGACTTCTGGCACACAAAGGCTCTCCCAAGCATGGAGTCCCCTCT  458
          |||||||||  |||||||||||||| ||||||| |||| |||||| |||||||||||
Db  8557  CTGCTTTCCGGCACTGACCTTTGCCATACTGACCAATTCCTCGTCTCTTAACATTCCTTCC  8616

Qy  459   CTCCGCTTTACAGATGGCCCCACCGGTAAGAGGGACCAAGTTGTTCTTCAAATGCGGTCCCT  518
          |||||||||| ||| ||||||||||||| ||||||||||||||||| ||||||||||
Db  8617  ATTAGATTATCAGACGGCCCTAATGGTATCCGTGGAACTAGCTTCTTCAAATTCTTCACCA  8676

Qy  519   GCGGCCTGCTGTTCCCTTGCGCGCCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAA  578
          |||||| ||||||||||||||||||||||||||| ||||| ||| ||||||||||||
Db  8677  TCCCGTTGTTTTCCCTGTGGGACAGCGCTAGGCGCTACTTCGACAAAAAGTTACTTACTTC  8736

Qy  579   GAGGCAGGTAAGANGATGGGCAAAGAGAGGGCCATGCGCTAAGAGTGGCANGTGATCCTCGGC  638
          ||||||||||||| |||  |||||||||||||||||| ||||||||||||||  ||||
Db  8737  GAAGTTGGTGAAATATTACCAGAAGAACCGAAAGGTGTTAGTGTGGTTTGGGT  8796

Qy  639   CCGACTATCAACATGCAAGCCTCCCCTCTCCGGTGGACGTGGCTTCGAGTGCGATGGGTGAG  698
          ||||||||||||||||||||||||||| |||||||||||||| ||||||||||
Db  8797  CCAACGGTAAAATATCCATCGTGAGGTTAGGGGTTTTGAATCATTTTCTGAA  8856

Qy  699   GATCCGGTTCCTGCGGGCGGCTTGGGAGCTCCGGGCTCTCATCCCGCGGGGGCATTCAGAGCACTGGA  758
```

```
Qy   1827 GAGAGGGGCCGGCAACTCTGTCAAGGGCAACACGTACAAGTTCAAGATCGAGTTCGGC 1886
               ||||  ||| || ||||  |||||
Db   9994 GAGAGAAATTCTATTTATTTATTAAAGAAGGGAGAAAGTATAAAGTACGGGTCGAGTATGGA 10053

Qy   1887 TCCGGCACCACTACACCCTCAAGGGCGACACCAACGTCCCCGGCCACGGCTCCTCCGC 1946
               |||||||
Db  10054 AGTGCAGCCACTTACACGCTGTCAACTAATCTGTCTCCCAGTACGGGAGAAGATACTCT 10113

Qy   1947 GTCGGCGGCTGCCAAGGTCATTGACGACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCC 2006
               | |||| ||  |||  |||||| |||||  |||| |||   ||||||| |||
Db  10114 ATTGGCTGTCGTGTTAAAGTTAATTGATCCTGAAAACGATAGATTACGCAGTTGAGGGTTGCC 10173

Qy   2007 AAGGACACCAGCCACCAGTTCATCATCAGTGCAGGGCCTTAACGCCGACTGGGAGAGGGC 2066
               | |||||   ||
Db  10174 AAATCAGTTGATTGTGTTGATTTTATTGTGTTGGAAACGAGAGTACGCAGTTGAGGAAGT 10233

Qy   2067 GCCGACGCCGGCGAGCATGAAGCTCCCCGGGGTGCTTGGGACCACCAACGTCCCCGGCC 2126
               | |||| |||  |||  |||  |||     |||| ||||| |||| |||  ||||
Db  10234 GAAGATAGAAAAACAATGACTTACCATCGGTTATCTGACAAACTGGTTTATTCTTATTCTTG 10293

Qy   2127 GCCGCGGCAACCCAAACACACGCTTCGTGCTCGTGTCATGCAGACGGGGCAGAGAGATCCCCTGG 2186
               | |||  |||      |||||  |||  |||    ||||||||| |||   ||||
Db  10294 CAATCCAATCCCAAATACAGTAGTTGTTGTGACTCAAATCAGGCACTCCAAATTGAGATCGCATGG 10353

Qy   2187 CTCGACGCCACGGCCGGCGTCATCCAGGCCGGTACGGCGGGACGAGACGGGCAACTCC 2246
               |||| |||||   |  | |||||||||  | ||| ||  ||| |||  |
Db  10354 ATCTCTGAAGCCCCACACCCTTATTGCACATTGGTACAAGGTAACGAGAATTGAAAATGCT 10413

Qy   2247 ATTGCGCACGTCGTCTTTGGCGACTACACACCCGTCACCAACGCTCATTGCCTCAGCTTCCCC 2306
               |  |||||       ||||   |||||     |||  ||||  |||| ||||||   |
Db  10414 TTAGCAAACATTATTTTTGGAGAGCAAAAACCCTTGTGGAAAAATTGCCAATAACATTCCCT 10473

Qy   2307 AAGCCCCTGCAGGAACAACACCCCGCGTTCTCCAACTTCCCGCCACCGGGCATCGCT 2366
               |  |||| |||   || ||  || |   |||  |||  ||  ||||||  |||   |||
Db  10474 AAAAAACTGAAGGATAATCCTGCCTANCTTTCATTTCGTTCGTGTGGCCATTNGTGTT 10533

Qy   2367 TACGGCGGAGGACGTCTACGTCGGGGTACAGGCGGTTGCCGACAAAGGACGTCAAT 2426
               |  ||| |||||||||  || |||||  |||  ||||  || |||||| ||| ||||
Db  10534 TATGGAAGAGGACAGTCTTTGTTGGGGTATAAGTTATTATGAACCCGTAGAAAGGAGGTCTTG 10593

Qy   2427 TTCCCCTTTGGGCCACGGCGCCTGTCCTACACACCACTTTTGCCTTTCCAATCTCTCCGGTGTCT 2486
               |||   |||  || || |||||||  || ||||||      ||||  |||||  ||| |
Db  10594 TTTCATTTGGATATGGGCGTTCATATNTACTACTTTTGAATTATNCCAATNCTTTATTTGAAG 10653

Qy   2487 CACAAGGACGGCAAGTGTGAGGGTGTCCCTCGCGGTGTCCTTCCGGTGAAAGAACACACGGCCT 2546
               ||  |||| || |||  ||  || |||||||||||| ||  || |||| ||| |||| |
Db  10654 AATTGTGGTGGAAAGGTTAAGGACGTTGCTTCATATACTGGCCCCCATGTGTCTGGT 10713

Qy   2547 GCACAGGTGGCCCAGCTCTACGTCAAGGCCCTTCCAAGCGGGGACATTTGGATGAGGCGGGATT 2606
               |  ||| ||  ||||| ||   ||| |||  |  ||||  |  ||| |||| |
Db  10714 GGATTTTCAGTTTTTCATTTGAATTATCCAAAATCTTTATTTGAAG 10770

Qy   2607 AAGGAGCTCAAGGGCTTCGCAAAGGTCGCAACGTCGACGGAGCGAGACGAGGAAGCGCGGTGGCGACA 2666
               || ||||| |||  |  | |||  || ||| ||  ||| | ||  |||| |||||  |||| |||
Db  10771 AAAGAACTAAAAGAGTTTTCTAAAGTCGTTTTATGGTCCTAAAGAGACCAAATTAATTAGA 10830

Qy   2667 ATCGAGGAGCAGGAGAAGTAGGTAGTACGGATTAGAGAGGCGGGATCAGTGGTGCT 2726
               ||  |||||| ||  || |||||
Db  10831 AAAGAACTAAAAGAGTTTTCTAAAGTCGTTTTATGGTCCTAAAGAGACCAAATTAATTAGA 10890

Qy   2727 GTCGGAAAAGGGTGACTATGAGGTTATCGTGACGGACGCACGCGCAG 2772
               ||  ||||  ||||||| || |
Db  10891 TCTGAAGAGGGGGAGTACAACGTTCTTGTTCGGAAACGAGCAGTAAAG 10936
```

```
RESULT 4

LOCUS       SPBC1683        41799 bp    DNA     linear   PLN 23-AUG-2000
DEFINITION  Schizosaccharomyces pombe cosmids c1683.
ACCESSION   AL355920
VERSION     AL355920.1  GI:7801293
KEYWORDS    adenosine deaminase; alpha-glucosidase; beta-glucosidase precursor
            (EC 3.2.1.21); ferric reductase transmembrane component; frp1;
            ght4; glucose transporter protein; glycosyl hydrolase family 3;
            inosine-uridine preferring nucleoside hydrolase; isocitrate lyase
            (EC 4.1.3.1); MFS allantoate permease; MFS inorganic phosphate
            transporter; TFl-type LTR; transcriptional activator; zinc finger.
SOURCE      Schizosaccharomyces pombe
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces;
REFERENCE   1  (bases 1 to 41799)
  AUTHORS   McCombie,W.R.
  TITLE     Sequence analysis of a region of the fission yeast genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 41799)
  AUTHORS   McCombie,W.R. and Lyne,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) W. Richard McCombie, Cold Spring Harbor
            Lab, 1 Bungtown Rd, P.O. Box 100, Cold Spring Harbor, NY 11754, USA
            and Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            CDS are numbered using the following system eg SPAC5H10.01c (S.
            pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid SPBC1683 is overlapped at its 5' end by cosmid SPBC359 and
            at its 3' end by cosmid SPBC1198.
FEATURES             Location/Qualifiers
     source          1..41799
                     /organism="Schizosaccharomyces pombe"
                     /db_xref="taxon:4896"
                     /chromosome="II"
     misc_feature    1..500
                     /note="nominal overlap with cosmid SPBC359 S. pombe
                     chromosome II"
     mRNA            1351..1543
                     /note="mRNA from spc08587 1 193"
                     /map="IIL"
                     /clone="cosmids 359 (partial), 1198 and 1683"
                     /clone_lib="Mizukami, et al., 1993. Cell 73;121-132."
     gene            3226..4947
                     /gene="SPBC1683.01"
     CDS             3226..4947
                     /gene="SPBC1683.01"
                     /label=SPBC1683.01
                     /codon_start=1
                     /product="MFS inorganic phosphate transporter"
                     /protein_id="CAB91163.1"
                     /db_xref="GI:7801294"
                     /note="SPBC1683.01, len:573,
                     SIMILARITY:Schizosaccharomyces pombe, YAEC_SCHPO, putative
                     inorganic phosphate transporter c23d3.12., (559 aa)," fasta
                     scores: opt: 2395, E(): 0, (70.0% identity in 574 aa)"
                     /translation="MKLNIFKSHGDSNTAEERPVPLEQVEAEDQOHENRFWLGITAKE
                     FRLMLAVGFFLDSYDLFIINLVPIFEYLIYWGGIEKGPNGKGHYPSGIRGLVNAAS
                     NIGNIFGOLMFGFMGDFGRKFVVGKEMIIVIATAIILLIAMPKSIHSPLSKMMVFCW
                     RWLGVGIGGDYPMSAAITSERSKLNRRGTILSLIFAFQGFGTLAGAIVTIILLGCFE
                     HPLNREGHYRKLEGYWRLQPGLALVPAIGVLIPRLMMEETQKFNSQOLNSGGNDRPK
                     TSLNFEDDELVKNPSVTKGHPEIHESSENYLSRSNTVENEPENTEKQOFESVSAPANRS
                     GFIQYFPQWHHFKHLLGTSVCWFLLDIAFYGVNSVILKNIGFSSGTNEYRTLMKN
                     AIGNIIAVAGYVPGYWFNVFLVELIGRKWIQLQOFVITGLMFALLAGRWNEISGGR
                     FACFVIAQLFSNFGPNSTFIYPAEVFPARVRGTPAHGISAALGKCGAILASLLENFLT
                     SIIGYGNVMWIFCGCWGGILFTLLLLPETRGRDADEIDRVELFYGGDGVECNSKWKS
                     WYVNGIF"
```

```
misc_feature    join(3364..4004,4134..4872)
                /gene="SPBC1683.01"
                /note="Match to PF00083 sugar_tr, Sugar (and other)
                transporter Score 132.74"

LTR             complement(5308..5478)
                /note="partial degraded TF1-type LTR"

gene            5967..6986
                /gene="SPBC1683.02"

CDS             5967..6986
                /gene="SPBC1683.02"
                /note="SPBC1683.02, len:339, SIMILARITY:Saccharomyces
                cerevisiae, ADA,YEAST, Probable adenosine deaminase, (347
                aa), fasta scores: opt: 723, E():0, (37.6% identity in 351
                aa)"
                /codon_start=1
                /label=SPBC1683.02
                /product="putative adenosine deaminase"
                /protein_id="CAB91164.1"
                /db_xref="GI:7801295"
                /translation="MTIERFIEKLPKAELHLHLEGTLEAEIKLKLSHRNKIPLKQSS
                IEETKESTNFHDLASFLEVYYEGVEILLHEQDYDLCYQILRKAASONVVAEMFFDP
                QLHTRRGISFPETVIKGIIRARDDAMRDHIIYSOLIMCFIREMSFENAEETLNASLPYK
                SEIIGIGDSNEENNPPIKFLKVFQRARQLGRLTCHCDLHQKNTTHIRQALEDIGV
                ERIDHGINILDDPELIKLALERINIPFTVCPFSNEIVPGKAQPEIRIMLDTGLKVTIN
                SDDPAYMHCFYITENFNLAQKGASLTKKELVQICRNSPEAANISEEKRNHVLEALNSF
                ASAYD"

misc_feature    6141..6956
                /gene="SPBC1683.02"
                /note="Match to PF00962 A_deaminase, Adenosine/AMP
                deaminase Score 164.32"

gene            complement(7448..9007)
                /gene="SPBC1683.03c"

CDS             complement(7448..9007)
                /gene="SPBC1683.03c"
                /note="SPBC1683.03c, len:519, SIMILARITY:Saccharomyces
                cerevisiae, Q08902, orf yor378w, (515 aa), fasta scores:
                opt: 2052, E():0, (58.0% identity in 509 aa)"
                /codon_start=1
                /label=SPBC1683.03c
                /product="MFS transporter of unkown specificity"
                /protein_id="CAB91165.1"
                /db_xref="GI:7801296"
                /translation="MMNVVKSLGTYNTNMEKRSSKDLVQNEDTPLPVQKISMSLFHE
                IVFFIACTAQLMTQAGLGQSIAPNNIIGKSLGTTNPGOLSWFPASYSLTVGTFILIA
                GRLGDIYGHKMFVLGIYIMFCIWSLISGSYSYAKSVIMFDVCRALTGIAPAFLPNAL
                ALGRVYPPGKKKNLIFIALFGATAPNGLELLGSVFGVSGLVLINFSWNQGPVVGWQVPVYVILF
                IIGELSIVVFLVLEKRVVOPILAPSMNNSEMGCVLICVAAGWACFGIKMYYLMQFLEN
                LRFAPLLVTAQLTPVGISGCAAALTTGYLRLKPLRKNVFMIVYVSMIAFIVYGTILIIATAP
                IHQTYWAQFVSIIVTPWGMDMSPPAAILMLSDEVPKQHQGIRASLVSTVVNYSISIG
                LGIAGTVEHLNHKGVDLIRGYRSAWYMGTGFGGLGVCVAILYFASYLKVGQQKSPK
                FPLIMKNTKAV"

gene            complement(20737..21669)
                /gene="SPBC1683.06c"

CDS             complement(20737..21669)
                /gene="SPBC1683.06c"
                /note="SPBC1683.06c, len:310, SIMILARITY:Escherichia
                coli, YBEK_ECOLI, hypothetical 33.8 kd protein in
                leus-gltI intergenic region, (311 aa), fasta scores: opt:
                838, E():0, (44.5% identity in 308 aa)"
                /codon_start=1
                /label=SPBC1683.06c
                /product="putative inosine-uridine preferring nucleoside
                hydrolase"
                /translation="MEDPKDSDEKFDIGISEKNLDVGFGESSSSVDVPVKGRFASFLKKL
                ELSSGPEKENIDLRPTPPDRHYSALDIIYLWSCNGISASAFRTGTSYMEMGLSPKQA
                LAALIAGNVFIAMPMTLNGLRGSHYHIPFAVQSRASEGYYFWNLILLRFIAGLFYYG
                TNYTTGACVQTLIYAIFKSFRSYKNRLPADAGITSDELISYFYWVISFPHLIRPE
                YLORFFLIKSISTYIACFAMLIFLICNVGSHVVWDQPATVSGRSWSWVFPMCALNSSVA
                GFSTLAVVNVNDFTRYKVHKPKTPVVOMLILPLVAAVSAPIGIVSGVASKIMYGTAMWDP
                LQIANNWTSRGRRAAAFFMGLTVIVSMIRAQNISDNTVAAANDLLYFFPRYLDRRQV
                PSGIYRYNKLGLNWRALLIAFLCACVPLIPGMAMSINPSITMPDGVIHLYYIGYFYSFM
                TAFIIYWGLNLVFPAKETLLEEAVYPRKSNAELVDPSTLSGKDKFWIYIDY"

gene            17673..19352
                /gene="SPBC1683.05"

mRNA            17673..19352
                /note="mRNA from spc06472 118 1"

CDS             17673..19352
                /gene="SPBC1683.05"
                /note="SPBC1683.05, len:559, SIMILARITY:Saccharomyces
                cerevisiae, THIY_YEAST, putative thiamine transporter
                yor192c., (599 aa), fasta scores: opt: 828, E():0, (26.8%
                identity in 512 aa)"
                /codon_start=1
                /label=SPBC1683.05
                /product="NCS1 allantoate transporter"
                /protein_id="CAB91167.1"
                /db_xref="GI:7801298"
                /translation="MEDPKDSDEKFDIGISEKNLDVGFGESSSSVDVPVKGRFASFLKKL..."

misc_feature    11428..13776
                /gene="SPBC1683.04"
                /note="Match to PF00933 Glyco_hydro_3, Glycosyl hydrolase
                family 3 Score 687.70"

gene            11335..13833
                /gene="SPBC1683.04"

CDS             11335..13833
                /gene="SPBC1683.04"
                /note="SPBC1683.04, len:832, SIMILARITY:Kluyveromyces
                marxianus, BGLS_KLUMA, beta-glucosidase precursor, (845
                aa), fasta scores: opt: 1274, E():0, (46.3% identity in
                848 aa)"
                /codon_start=1
                /label=SPBC1683.04
                /product="beta-glucosidase precursor (EC 3.2.1.21);
                glycosyl hydrolase family 3"
                /protein_id="CAB91166.1"
                /db_xref="GI:7801297"
                /translation="MMEHDVEDLINQLDISEKAMLLSGTDLWHTAAIPRLNIPSIRLS
                DGPNGIRGTSFFNSSPSACPPCGTAIGATFDKKLLFEVCGEYLAEAKAKGVSVVLGPT
                VNIHRGPLNGRGFESFSEDSTLSGLAASTVILGLQOSKNVQACIKHFVCNDMEDERNSV
                SIDVSQRALREVYLMPFQLACKYSNFKKSLMTSVKVNGHEVSQSRILLDNILRKEWEW
                KGTIISDWFGYSYKAIDAGLDLEMPGKPRFRNVNTIQHLVGSKELSESIIDERAKN
                VLKIVKHSWQNTEAEHCELNNDSSCLREALKFASQSIVLLKNKKKLLPLSRKGTFA
                VIGPNAKVCNYSGGGSANLKPYIVSMYDGIAAKIDGVPEYALGCHNNILNPIANLL"
```

```
Query Match
Best Local Similarity   13.2%;         Score 393.6;  DB 8;   Length 41799;
                        49.0%;         Pred. No. 2.1e-42;
Matches 1199;  Conservative  0;  Mismatches 1229;  Indels  18;  Gaps  5;

Qy    339 ATGGCTGAATTGAGATGTTGAGGCCATCTTGAAGAAGCTCACCTGGCCGAGAAGGTCGAT   398
          |||| || |||||||| |||||||||| ||| ||||| |||||||| | ||| ||||||
Db  11335 ATGATGGAACATGATGTTGAAGAGATTTGATCAATCAATTGGATATAAAGTGAGAAGGCTATG 11394

Qy    399 CTGCTGCTGGCTGGTCGTATCGACTTCTGGCACACAAAAGGCTTCTCCCCAAGCATG      458
          |||| ||   ||   || |  ||||  ||||||||| ||| |||| |||| |||
Db  11395 CTGCTTCCGGCCACTTTGGCATACTGCCAGCAAAGTTCCTCCGTCTTAACATTCCTTCC   11454

Qy    459 CTCCGCTTACAGATGGCCCCAACGGCGTAAGAGCGACAAGTTCTTCAATGGGGTCCCT      518
          || || ||  |||| |||| ||  || ||||||    |||||||| ||||||||||||
Db  11455 ATTAGATTATCAGACGGCCCCTAATGTATCCGTGGAACTAAGCTTCTTCAATTCTCACCA   11514

Qy    519 GCGGCCTGGCTCGCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAA 578
          ||| |||  || ||  || |||| |||| |||  || || ||| ||  ||| ||| |||
Db  11515 TCCGCTTGTTTCCCTGTGGGACACCGCGCTAGGGGTCTACTTCGACAAAAAGTTACTATTC   11574

Qy    579 GAGGCAGGTAAGATGATGGGCAAAGAGGCCATCGCTTAAGAGTGCGCGCATGTGACCTCCGGC 638
          || |||||| ||   || ||| || || ||||   |||||||| |||  |||| || ||
Db  11575 GAAGTTGGTGAATATTTAGCAGAAGAAGCAAAAGCGAAAGGTGTTAGTGTTGGTTTTGGGT   11634
```

```
Qy    639  CCGACTATCAACAATGCAAGCNGCTCCCCTCTCGGTGGACGTGGCTTCGATTGGTTGGAG   698
           ||| ||||||  || | |||| ||||||| ||||| ||||||  |||||| |||| |||
Db  11635  CCAACGGTAAAATATCCATNCGTGGGACCCTTAAATGTGTAGGGGTTTGAATCATTTTCTGAA  11694

Qy    699  GATCCGGTTCCTCGGCGCGGCTTGGGAGCTGCGCGGCTCTCATCCCGGCANTTCAGACGCTGGA   758
           ||| |||  |||| ||| |||||||||||||||||| || | |||  ||||  ||||
Db  11695  GATTCTACTTTATCAGGTCTTCGAGCTAGTTANTTANTTCTTGGATTGCAAAGTAAAAAC  11754

Qy    759  GTGCAGGCTACGATCAAGCACTTTTGTGCAATGATCAGGACGCCGATGATGGTG   818
           ||  | |||||||||  ||  |||  |||||   ||||   ||||  ||||  ||
Db  11755  GTCCAACCATGCATTAAGCATTTTGTGTATTGATGATGAGAGATGAAAGGAAACTCGTC  11814

Qy    819  CAGAGCATCGTCACGGAGGAGGCGCTCTCCGTTGAAATCTACGGCACTCCCGTTCCGAGATTGCT   878
           || | |||| ||||||  | |||| ||||||| | |||||| ||||| || || ||  |||| ||
Db  11815  AGTATTGACGTCTCACAAAAGGCCACTTAGCAGAAGTATATCTTATGCCATTTCAGTTGGGCA  11874

Qy    879  GTCGGAGACTCCCAGCCGGGGTTGCGTTCATGACGGCANTCAANTGGCANCANTGACGCTGTCG   938
           ||||||| | |||||  |||| |  | || ||||||| |||||||  ||  | ||  ||||||
Db  11875  TGCAAATATTCAAATTCAAATCACTTATGACTTCTTACAACAAGGTGAAATGGTGAACAC  11934

Qy    939  TGCAGCGAGAACCCTAAAATATCTTACGGCACTCCCGTTCCGATTGCT   998
           ||||| ||||   ||| || ||||||||  |||  ||  || |||||
Db  11935  GTTTCCCAAATCTCGAATTTTATTAGACAAATATTTTAAGAAAATGGGGAGTGGAAAGGT  11994

Qy    999  CTAATCATGAGCGACTGGTACGGCACATACAACAGTACCACACAGAAGCCGTTGTGGCAGGCCTC  1058
           ||| |||  | ||  | ||||| ||  ||  ||||| ||||||||| |||||||||||||| |||
Db  11995  ACTATAATATCTGAATTGGTTTCGGGAACTTAATCGTTGAAGAAAGCTATAGATGCAGGTTTG  12054

Qy   1059  GACCTCGAGATGCCCGGACCTCCACCGCTTCCACCGCTTCCGAGGAG---AAACACTCAAGTTCAACGTC  1115
           |||| ||||||||  |  ||||||||||  || | |||||||||||   ||||| || |||||||
Db  12055  GATTTGGAGAGATGCCAGGAAAGCCGAGAGTTCAACACTTAGTAGTCAACAC  12114

Qy   1116  TCCAACGGCAGAAGCCCTTTTATCCACGTCATTCGACCACAGAGGGCCTAGGGAAGTTCTTCAGTTC  1175
           ||| |||  ||| ||| | ||| |||| || ||||| ||||   ||||| ||||   | |||| ||
Db  12115  GGATCCAAGTAAACGTTTCGGAGTCAAATTTTAGATGAAGAACGCACCAAAGGACTGCCCTGAAGCTT  12174

Qy   1176  GTCAAGAAGTGTGCTCGCCTCCGGAGTGACGGAGAAGGTTGGCAACGAGGTCGGTGCTGCTGAAG  1235
           ||||||| || |  || |  |||||| |  | |||||||||||| |  | || | |||| ||||
Db  12175  GTGAAACATTCCTGGCGAATACAGAGCCGCAAAATCATTGTGAGCTAAATAATGATTCC  12234

Qy   1236  ACCCCCGAAAACGGCAGAGCTCTCCTCCGGAAGGTTGGCAACGAGGTTGGCAACGATTGGGCAGTGAAG  1295
           || ||||  |||| |||| |||| |||| ||  ||||||| |  |  ||||  ||||| |||| ||||
Db  12235  TCTTGTTTAAGGAGAGC---ACTTAAAAAAATTTGCAAGGCATCCAGTCTCATTGTTCTTGAAA  12291

Qy   1296  AACGAGAACAACGTTCTGCCCTTGACCAAGAAGAAGAGAAGAGGTCCCAGTACTAGTTGTCGGCCCCAAC  1355
           ||||| || ||||  || ||| ||||||||| ||||  ||||| |||| ||||||| |  ||||| |||
Db  12292  AATAAAAAGAAAACTTTTGCCATTGCAAGGAAAGGGCACGTTTGCAGTTATTGGAACCTAAC  12351

Qy   1356  GCCAAGCAGGCCACATAACCACGCGGCGGAGGCTCTCGCCGGCCTACTACGGCAGTC  1415
           |||| || ||| |||| ||||| |||||||||| | | || || | |  ||||||
Db  12352  GCAAAGTTTTGTAATTACACGGGATCGGCAAAATTTGAACCTTATTATACTGTG  12411

Qy   1416  ACTCCCCTTTGACGGCCTCAGCAAGCAGCAGCGTCCGAGACGGCCCATCGTACACCGTCGGGCGCC  1475
           || |||||| ||| |||| |||| || |  ||||| |||||| || |||||||||| |||||||||
Db  12412  AGTATGTATGCATGGCATAGCGACGGTGCCGCAAGGTAGCGTGTCTTGGTTGGC  12471

Qy   1476  TACACC------ACCGTTCCTCCCANTTCTAGGCGAGCAGCAGTGCCTCAGCGGCCTGCGGCGCT  1529
           |||| |      ||||||||| |   || | |||  | || || || ||| |||||
Db  12472  CATAACTATCTANAACCTTCCAAANATTATTGCAAACCTAGAACCGGCAAG  12531

Qy   1530  CCGGGCATGCGCTGTCGAGGGTCTTCAACGAGCCCCCTGGTACCCGTAAACGCCCCAGCACATT  1589
           ||||| ||| |||| |||| ||| |||| ||| |||| |||||| | || ||||| |||| ||
Db  12532  CATGGATACGTTGCCGAAGTTTTTATCTTGAACCTGCAAACCTTAGAACCTTGTGATA  12591

Qy   1590  GACGAGGCTCTTCTCACCAAGACGGACACATGCGAAGGTTGGGACTACTACCACCCCAA---G  1646
           ||||||| || | |||| || |||| | ||| | |  |||| |||||||||| ||||||   |
Db  12592  GACGATTACGATTTAGAAGATGGAGTTGTTCGTTTTTACGATTATTGTGAACGACAAAAATG  12651

Qy   1647  GCGGCAGAGACACGTGGTACCGCCGACGCAGCGAGGACTGCACC  1706
           || |||||| || |  |||| ||| ||| |||||| |||| || |
Db  12652  AAGGATGGGCTATTTCTATATAGAGACATTGAAGGTTATTATTT  12711
```

```
Qy   1707  TACGAGCTCGCGCCTCGTCCGTCGCGCTGCGCGGCAGCGGCAAAAGGCGTACGTAGACGACCAGCTCGTC  1766
           || || |  ||  |  | || |  ||  ||||| ||||| |||||  ||  |||  |||  | ||||||
Db  12712  TATGAATTGGGATATCTGTGTTGTTGGGAACGGCGCTCGTGCTTATTGATGATGTCTTATTA  12771

Qy   1767  GTCGGACAAACGCCACCAAGCAGCAGGTCCCCGGCGATGCCTCTCCGTTGATTAGGCTGTGGA  1826
           |||| | |  ||  ||| ||  |||   ||||  |  | |||  ||||| |  || || | ||
Db  12772  ATCGACAACAGAGCAAAAGCCAAACCCCTACGAAATCATACTTTTCTCTGTTTGGAATTGAA  12831

Qy   1827  GAGACGGCGGCCGATCAAATCTCGTCAAGGGCAACACGTACAAGATCGAGGTCGGC  1886
           |||| || ||  || ||| || ||  ||||  |||| ||| | ||  ||| |||
Db  12832  GAGAGAAATTCTATTTATTTTATTTAAAGAAGGGGAAAAGTATAATGTACGGGTATGGA  12891

Qy   1887  TCCGCACCCACCTACACCCTCAAAGGGCGACACCATCGTGCCCGGCGGCACGGGTTGCC  1946
           |||| |  |||||| |||| ||||| | || || ||||| | || ||  | ||| |||
Db  12892  AGTGCAGCCACTTACACGGTTCAAGTCGCTCAACTAATCGTCTCCCAGTACGGCGGAGGAGTATT  12951

Qy   1947  GTCGGGCGGCTCGCAAGGTCATTGACGGCATCAAGGGGCAACGGTGAAAATGTCCGTGCCCTCGC  2006
           || || | || ||| |||||| ||||| ||  |||||| || | |  |||| || |||||| |||
Db  12952  ATTGGCTGTGTTGTAAAAGTTTGGTTATTGAATCCTGAAAACGGAGATAGTAATGTACGGGTTGCC  13011

Qy   2007  AAGGAGCACGACCAGGTCATCATCCTGCGCGGGGGCCTTAACGCCGGCGGGAAGAACGAGGGC  2066
           || | | | | || | ||  | ||  |  |||| |||| |||| ||| |||   |  || |
Db  13012  AAATCAGTTGATTGTGTGATTTTTAGTGGTTTTGACACCGAGAATGGGAAACGGGAAGGT  13071

Qy   2067  GCCGACCGGCGGAGCAATGAAAGCTCCCCGGGGGTGCGTGGACGTCATTGCCGAGGTGGCC  2126
           | | | |||| | ||| | ||||| | ||| | | | |||| | || ||  |||| |||
Db  13072  GAAGATAGAAAAACAATGACTTTACCATCGTTATCTGACAAAACTGGTTTATTCATCG  13131

Qy   2127  GCCGGCAACCCAAAACACGTGTCGTCGTCATCGCAGACCAGCGCGAAATCGAAAGTCCCGGCCCTGG  2186
           ||| | || ||||  || |||  ||  | || || ||| ||   ||||  ||||| |||| ||||
Db  13132  CAATCCAATCCCAAATACAGTAGTTGTGACTCAAATCAAGGCGCACTCCAATTGCCATGG  13191

Qy   2187  CTCGACGCCACACGCCCGCCGTCATCCAGGCCGTCTGGTACGGGCAACGGTGACGGGCAAACTCC  2246
           || |  ||| ||  |  || | || ||| || || ||||| |||||||||| || |  | |||
Db  13192  AATCCTGAAGCCCACACCTATTTGCACATTTGGTACAACGGTAACAGGTGAGGGGTTGCC  13251

Qy   2247  ATTGCCGACGTCGTCTTTGGGCCGACTACAACCCCGCGGGGGGCAAAGCTGTCCCTCAGTTCCC  2306
           || || |  | || |  |||| || ||||| || || |||| |||||| ||  |||||||| |
Db  13252  TTAGCAAACATTATTTTTGGAGCAAAACCCTTGTGGGAAAATTGCCAATAAACATTCCCT  13311

Qy   2307  AAGCGCCTGCAGGACAAACCCCGCGTTTCTCAACTTCCGCACCGACGGCTG  2366
           ||| | ||| || || | |||  |  | |  |  ||||| |||| | |||| |
Db  13312  AAAAAACTGAAGGATAAATCCTGCCTATCTTTCATTCGTTCAAGCCGCCATTGTGTT  13371

Qy   2367  TACGGCGAGGACGTCTTACGTCGGTGGTACTACGACGTTTGGCGACAAGGACGTCAAT  2426
           ||| |  | | ||  ||||  |||||||||| ||  ||||||| ||  || || ||||
Db  13372  TATGGAGAAGACGTCTTTTGTTGGGGTTAACAGTATGAAGCCGTAGAAAGGAGGTCTTG  13431

Qy   2427  TTCCCCTTTGGCCACGGGCCTGTCCTACACCACTTTGCCTTTTCCAAATCTCCGTGTCT  2486
           ||| | || || ||| | || | | |||| |  ||| || || ||  | | |||| ||
Db  13432  TTTCCATTTGGATATGGGTTTTCATAATACTACTTTTCCAAATCTTTATTTGAAG  13491

Qy   2487  CACAAGGACGGCAAGCTGACGGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGGCGGC  2546
           ||  || | ||||| |||| |  |||||| | ||||| | ||||| || || |||| |
Db  13492  AATTGTGGTGAAAGGTTAAGGAGATCGATTTCGAAATACCAGTCTCGGT  13551

Qy   2547  GCACAGGTGCCCAGCTCTCACGTCAAGCCCCTCCAAGGGCCCCGTC  2606
           |||| | | | || | ||||  |||||| |  || |||  |||||
Db  13552  GCAGAAATTATACAAGTATATAATTTCTCAAATC---GTTCGTTCAGTTAATCGACCTGTT  13608

Qy   2607  AAGGAGCTCAAGGGCGGCTTCGCGAAAGGTCGAACTGCAGCCGGACGCAAGGGGGTGACA  2666
           ||||| ||| ||| ||| |||| |||  || ||| |||| || | ||||| ||  ||||
Db  13609  AAAGAACTAAAAAGAGTTTTCTTAAAGTGGTTTTTAAGTCCTAAAAGAGACCAAATTAATTGA  13668

Qy   2667  ATCGAGGAGGAGGAAGTAGCTACGTCGCGCTGGGTATTTTGATGAGGAGAGGGGGGTGTGT  2726
           |||  ||| ||  | || | |||| | || | | |||  | || | || ||||| || ||
Db  13669  ATTGAGTTGGATATAAAGTATGCTACTTCTTTTTTTATGATGAACTTAATGAAAAAATGGTGT  13728

Qy   2727  GTCAAAAAGGTGACTATGAGGTTATCCTGAGCGACAGCACCGCCAG  2772
           |||||| ||| | |||| | |  ||| | ||| ||| ||||| | |
Db  13729  TCTGAAGAGGGCGAGTACAACGTTCTTGTCGGAACCAGCAGTAAAG  13774
```

```
RESULT 5
SPU33010
LOCUS       SPU33010          85837 bp    DNA     linear   PLN 14-NOV-1995
DEFINITION  Schizosaccharomyces pombe cosmids 359, 1198 and 1683.
ACCESSION   U33010
VERSION     U33010.1  GI:1063617
KEYWORDS    .
SOURCE      Schizosaccharomyces pombe
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 85837)
  AUTHORS   McCombie,W.R.
  TITLE     Sequence analysis of a region of the fission yeast genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 85837)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1995) W. Richard McCombie, Cold Spring Harbor
            Lab, 1 Bungtown Rd, P.O. Box 100, Cold Spring Harbor, NY 11754, USA
FEATURES             Location/Qualifiers
     source          1..85837
                     /organism="Schizosaccharomyces pombe"
                     /db_xref="taxon:4896"
                     /chromosome="2"
                     /clone="cosmids 359 (partial), 1198 and 1683"
                     /clone_lib="Mizukami, et al., 1993. Cell 73:121-132."
BASE COUNT    27560 a   14781 c   15395 g   28101 t
ORIGIN

Query Match    13.2%;  Score 393.6;  DB 8;  Length 85837;
Best Local Similarity   49.0%;  Pred. No. 1.9e-42;
Matches 1199;  Conservative   0;  Mismatches 1229;  Indels  18;  Gaps  5;
```

```
Qy    339 ATGGCTGATATTGATTGTTGAGGCGCCATCTTGTGAAGAAGCTCACCCTGGCCGAGAAGGTCGAT 398
              |||  |||
Db  29116 ATGATGGAACATCATTGTTGAACAGATTTGATCAATTCAATTGGAATAATAAGTGAGAAGGCTATG 29175

Qy    399 CTGCTTGGCTGGTATCGCACTTTCTCGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCCTCT 458
          ||||
Db  29176 CTGCTTTCCGGCACTGACCTTTGGCATACTGCGAGCAATTCCTCGTCTTCTTAACATTCCTTCC 29235

Qy    459 CTCCGCGTTTACACGATTGGCCCCAACGGGGTAAGGGACCAAGTTCTTCAATGGGTGCCT 518
             |||||   ||
Db  29236 ATTAGATTATCAGACGGGCCCTAATGGTAATCCGTGGGAACTAGCTTCTTCAATTCTTCACCA 29295

Qy    519 GCGGCCTCGCTCCCTTGCCGGGCACGTCGCCTCCGGTTCCACATTCAACCAAAACTCTGCTCGAA 578
             |||||
Db  29296 TCCGCCTGTTGTTTTCCCTGTGTGGGACAGCGCTAGGGGGCTACTTCACAAAAAGTTTACTATTC 29355

Qy    579 GAGGCAGGTAAGATGAGTGGGCAAAAGAGGCCATCGCCTAAGAGTCCGCATGGATCCTCGGC 638
          ||||
Db  29356 GAAGTTGGTGAAATATTTAGCAGAAGAAGAGCAAAAGCGAAAAGGTGTTAGTGTGGTTTGGGT 29415

Qy    639 CCGACTATCAAACAATGCAACAGCGCTCCCCTCTCGGTGTGGACGTGGCGTTCGAGTCGATTGGGTGAG 698
          ||||
Db  29416 CCAACGGTAAATATCCATCGTGGACCCCTTAAATGGTAGGGGTTTTTGAATCAATTTTCTGAA 29475

Qy    699 GATCCGTTCCTGGCGGCGCTTGGGAGCTCGGGCTCTCATCCTCCGGCATTCAGAGCCACTGGA 758
          ||||
Db  29476 GAATCTACTTTTATCGAGGTCTTGCAGCTTATTGTTATTTCTTGGATTGCAAAGTAAAAAC 29535

Qy    759 GTGCAGGCTACGATCAAGCACTTTTTGTTGTCAAATGATCAGGAGGACAGGCCCATTGATTGGTG 818
          ||||
Db  29536 GTCCAAGCATGCATTTTGTGTGTAATGAAGATGAAAGGAACTCTGTC 29595

Qy    819 CAGAGCACATCGTCACGGAGCGGCGGCTCTCCCGGTCTGAAATCTACGCACTCCCGTTCCAGATTGCT 878
             |||||
Db  29596 AGTATTGACGTCTCACAAAAGGCACTTAGAGAAGTAATATTCTTATGCCATTTCAGTTGGCA 29655

Qy    879 GTGCGGAGACTCCCAGCCGCGGGTGCGTTCATGACGCGGCGACTGTCCGTACGTCTCG 938
             ||||| |
Db  29656 TGCAAATATTCAAAATTCAAATCACTTATGACTTCTTACAACGAGGTGAATGGTGAAACAC 29715

Qy    939 TGCAGCGCGAGAACCCTAAATATCTTGATGGGATTGCTTCGAAAGGAATGGGGTTGGGATGGC 998
             ||||
Db  29716 GTTTCCCAAATCTGAAATTTATTAGGACAAATATTTTAAGAAAAGAGAGTGGGAAAGGT 29775

Qy    999 CTAATCATTGACGGACTGGGTACGGACGTGCGACATAACAGTACGACAGAAGCCGTTGTCGGCAGGCCTC 1058
          ||||
Db  29776 ACTATAATATCTGATGGTTTTTGGGAAGTTATTCGTTGAAGAAAAGGTTCAACGTC 29835

Db  29836 GATTTGGAGATGCCAGGAAAGCCGAGAAATGTCAAACACTATTCAACACTTAGTA 29895

Qy   1059 GTCAAGAAGTGTGCCGACTCCACCGCTTCCGAGGAG---AAAACACTCAACGGTAAGTC 1115
          ||||
Db  29896 GGATCCAAAGAGACCTTTCGGGAGTTCAATTTTAGTAGACAATATTTAAGAAAAGAGAGCTT 29955

Qy   1116 TCCAACGGAAAGCCCTTTATCCCACGTCAGTTGACCAGAGGGGCTAGGGAAGTTCTTCAGGTTG 1175
          ||||
Db  29956 GGATCCAAAGAGACTTTCGGGAGTCAATTTTAGTGAAAGGCAAAAGAAGAATTGTCTGAAAGCTT 30015

Qy   1176 GTCAAGAAGTGTGCCTGCCTCCGGAGTGACGGAGAACGGCCCGAGACGACTGTCAACAAC 1235
          ||||
Db  30016 GTGAAACATTCCTGGCAGAAATACAGAAGCAGAAAATCATTGTGAGCTAAAATAAATTCC 30072

Qy   1236 ACCCCGCGAAACGGCCAGCTCTCCTCCGGGAAGGTTGGCAACGAGGGCAGCTCGTGCTGTGGAAG 1295
          ||||
Db  30073 AATAAAAAAAAGAAAACTTTTTTGCCATTGTCAAGGAAGAAGGAATGTGCTGAAGCTTAAC 30132

Qy   1296 AACGAGAACAACGTTCTCGCCCTTGAGCAAGAAGAAGAGAGACGGCTAGGGAAGTTCTTCATTGT 1355
          ||||
Db  30133 TCTTTGTTTAAGAGAAAG---ACTTAAAAAATTTGCAAGCCACGTCCATTGTTCTCTTGAAA 30192

Qy   1356 GCCAAGGAGGCCACATACCACACGGCGGAGGCTCTGCCCCACTCAGGGCCTACTACGGAGTC 1415
          ||||
Db  30193 AGTATGTATGGAGCATAGCACGCAGCGTTGGCGCCGGTCACACGTCGTAGCGGTGCTTGGTTGGC 30252

Qy   1416 ACTCCCTTTGACGCGCTCAGCAAGCACGTGATGGACGCCATCGTCACGGCTCGGCGCCC 1475
          ||||
Db  30253 CATAACTATCTAAACCTTCCAAATATTGCAAATATTATTAGTCAACCCTAGAACGGGCAAG 30312

Qy   1476 TACACC------ACCGTTCCTCCCATCTCTCCAGAAATAGAGCACGCGAGGGTGCCGGCTT 1529
          ||||
Db  30313 CAATAACTATCTAAACCTTCCAAATATTGCAAATATTTAGTCAACCCTAGAACCGGCAAG 30372

Qy   1530 CCCGGGCCATGCCGTGGAGGGTGTCTTCAAGCAGCCCCCGGTACCCTCACCGGCAGCACATT 1589
          ||||
Db  30373 CATGGAATACGTTGCCGGAAGTTTTTATCTTGCAACCTCCGAAAAATAGAGACTTTGATA 30432

Qy   1590 GACGAGCGTCTTCTTCACCCAAGACGGACAATGCCACCTGGTGGTGGACTACTACCACCCCAA---G 1646
          ||||
Db  30433 GACGATTACGATTTAGAAGATGGAGTTGTTCCGTTTTTTACGGATTATTGTAAACGACAAAATG 30492

Qy   1647 GCCGGCGAGACACGTGGTACGGCCGGACACTGAAGGCGCCGACCGGGACCGCCGACGGAGGACTGCACC 1706
          ||||
Db  30493 AAGGATGGCTATTTCTATATATAGACATTGAAAGGTTATTTAAATTCCTGACGAGGACGCTGTT 30552

Qy   1707 TACGAGCTGCTCGGCCTCGTCGTGCTGGCACGGCAAAAGGGGTACGTTAGACGACCAGCTCGTC 1766
          ||||
Db  30553 TATGAAATTTGGGATATCTGTGTTTGGACACGGCTCTTTTTAGTTGATGAGTGTCTTATTA 30612

Qy   1767 GTCGGACAAGCCACCACCAAGCAGGTGTCCCGGCGATGCCCGGTCTCCGGCGAGAG 1826
          ||||
Db  30613 ATCGACACACAAGACAAAGCAAAACCCTACGAAATCATACTGTTTTTTGAATTTGGAACAATAGAA 30672

Qy   1827 GAGACGGGGCCCGCAATCAAATCGTGCAAGGGCCAACAGTACAAGTGTTCAAGATCGAGTTCGGC 1886
          ||||
Db  30673 GAGAGAAATTCATTATTTTATTTAAAGAAAGGGAGAAAAGTATATAAAGTCACGGGTCCAGTATGGA 30732

Qy   1887 TCCGCCACCCACCTACACCTCAAGGGCGACACCAATCTGCCCGTTCCCAGATTCGT 1946
          ||||
Db  30733 ATTGGCTGTGTGTTAAAGTTATTGATCCTGAAACGGACACTGAGGGTTCGC 30792

Qy   1947 GTCGGGCGGCGCTGCAAAGGTCATTGACGACCAGGCGAAAAATCAGAAAAAAGTTCGTCGCC-CGCC 2006
          ||||
Db  30733 ...
```

```
LOCUS       AF329731                2793 bp    mRNA    linear   PLN 28-JAN-2001
DEFINITION  Volvariella volvacea beta-glucosidase mRNA, complete cds.
ACCESSION   AF329731
VERSION     AF329731.1  GI:12584216
KEYWORDS    .
SOURCE      Volvariella volvacea
  ORGANISM  Volvariella volvacea
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Pluteaceae; Volvariella.
REFERENCE   1  (bases 1 to 2793)
  AUTHORS   Ding,S., Ge,W. and Buswell,J.A.
  TITLE     Molecular cloning, characterization and expression of a
            beta-glucosidase gene from the edible straw mushroom, Volvariella
            volvacea
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2793)
  AUTHORS   Ding,S., Ge,W. and Buswell,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-2000) Biology, The Chinese University of Hong
            Kong, Shatin, N.T., Hong Kong, China
FEATURES             Location/Qualifiers
     source          1..2793
                     /organism="Volvariella volvacea"
                     /strain="V14"
                     /db_xref="taxon:36659"
     CDS             41..2629
                     /EC_number="3.2.1.21"
                     /codon_start=1
                     /product="beta-glucosidase"
                     /protein_id="AAG59831.1"
                     /db_xref="GI:12584217"
                     /translation="MPPSDFAKANIDEIVEQLILDEAISLITAGVGFWHTHAIERLGVP
                     AVKVSDGNGIRGNHFFEMGTPAKCLPSSTALGATWDPEVEVGVLKLLAPEARLRAAS
                     LVLAPTSNIQRNPLGGRSPESFSSEDPYLSGIISASYVNGVQKGIGATIKHPVGNDKE
                     DDRQGYDSIISERALREIYLLPFMLTQKYAAPWAIMIYAYNRVNGVHVAEDPFLLKQVL
                     RNENKYKGLIMSDWFGMYSVDHGIKAGIDLEMPGINKNRTLDLVNRTIQARKLTPRDI
                     KDRARVVLELVKKCAQGRAPEILDGDGEERTVELESDKLLMRRIASESIVLLKNDNVLP
                     LKPEGGAIKKIAVVGGNAKAQVLSGGGSAALRASYFISPYDGIKAALEPHGVEVTPSE
                     GARAYKTLPFTLEWDLETEIGRGWIGTWHTHESDDSMTALDQPFIARLVDEIFRIFIS
                     TSYPKGITKRWTMRLKGYLIKPREKDTNEFGLITAAGRAKLWDGQLVIDNWTRQRGE
                     AFFGSGSQEETGVYLLRAGKKHEIYEVCNVRAPADGDEDEAIMDSNPGVRLGGAEVA
                     NADDLLSEAVKLASEADAVIAVVGLNADMETEGNDRRTLALPGRTDELVEKVAKVNSK
                     TVVTVQAGSAITLPWLDSVAAVVHAWYLGNATGDAIADVLFGKRQNPSGKLSLITPKRL
                     EDVPSHGIHFGSENGKVRYAEDLFVGYKHYHHNNIEPLFPFGFGLSYTTFSLSNOLSA
                     PVIDHATSSFSLTATLSITNTGPVTGSEVAQLTVSYPETSELTHAPLQLRAFKKVKDL
                     QPGETREVKLELDKYAVSSYWNDRYQTWAVENGEYEIKVGNSSVSKDLTLVQRFVVKDG
                     FEWRGI"
BASE COUNT      697 a    696 c    737 g    663 t
ORIGIN
```

```
RESULT 6
AF329731

  Query Match            13.2%;  Score 392.6;  DB 8;  Length: 2793;
  Best Local Similarity  50.2%;  Pred. No. 3.7e-42;
  Matches 1268;  Conservative    0;  Mismatches 1184;  Indels  75;  Gaps  9;

Qy   2007 AAGGAGCACGACCAGGTCATCATCTCCGCGCGGCCTTAACCGCCGACTGGGAGACCGAGGGC 2066
               ||   ||   ||   |  ||| ||     |  |      || |   |    ||| ||
Db  30793 AAATCAGTTGATTGTGTGGATTTAATGTGTGGGTTTGACACGAGAATGGGAAACGGAAGGT 30852

Qy   2067 GCCGACCGCGCGGCGCATGAAGCTCACCCCGGGCGTGCTGGACCACGCTCATTGCGCTGGGC 2126
               ||  |||    || ||   |  || |   ||||  || ||  ||  ||  ||
Db  30853 GAAGATAGAAAAACAATGACTTTACCATCGGTTATCTGACAAACTGGTTTTTTATTCTG 30912

Qy   2127 CCGGCGGAACCCAAAACACCGTCGTCGTCGTCATGCAGAGACCGAGGAGGAATGCCCTGG 2186
               |  ||    |    ||  |   ||  | ||| |  || | |  |||  |  ||
Db  30913 CAATCCAATCCAATACAGTAGTTGTGACTCAAATCAGGACACTCCAATTGAGATGGCCATGG 30972

Qy   2187 CTCGACGACCCACGCCGCGTCATCCAGGCCTGTACGGCGCGGCAACGAGAGAGGGCAAACTCC 2246
               |  ||  ||  |  ||  ||  |||| |     |     |   || |||    ||
Db  30973 ATCTCTGAAGCCCACACTTTAGGTACAAACGGTTGGTACGAACAGAATTAGGAAAATGCT 31032

Qy   2247 ATTGCGACGTCGTCTGTCTTTGGCGACTACACACCCCTCGGGGCAAGCGTGTCCCTCAGCTTCCCC 2306
               |||   | ||   || |  |   | || ||   ||| | |  ||  |||    |||
Db  31033 TTAGCAAACATTATTTTTTGGAGAGCAAAAACCCTTGTGGAAAATTCCCAATACAATTCCT 31092

Qy   2307 AAGCGGCCTGCAGGACAAACCCCCGGGTTTCTCAACTTCCTCACCAGCGGGCGCACGCTG 2366
               ||  | || |  ||  ||    |||  |  |  |  || | | ||| | || ||
Db  31093 AAAAAACTGAAGGAATAAATCCTGCCTATCTTCATTTCGTTCAAGCCGTGGCCATTGTGTT 31152

Qy   2367 TACGGCGGAGGACGTCGTTACACGGTACTACGAGTTTGCCGACAAGGACGGTCAAT 2426
               || |||  | | | |   |  |  || ||   || |  ||   |||||  || |
Db  31153 TATGGAGAAGACGTCTTTGTTGGGTATAAGTATTATTATGAAGGCGTAGAAAGGACGGTCTTG 31212

Qy   2427 TTCCCCTTTGGCCACGGCGCCGTGTCTCCTACACACCACTTTTGCCTTTTTCCAAATCTCTCCGTGTCT 2486
               | |  ||  ||| | | ||  | |      |  |  || || |  ||| |  ||| |   ||
Db  31213 TTTCCAATTTGGATATGGCTTTCATCATAACTACTACTTTGAATTATCCAAATCTTATTTGAAG 31272

Qy   2487 CACACAGGACGGCAAGCTGACCGTGTCCCCTCGGTGAAGAACACACCGGCGTCCGTGCCGGC 2546
               |  ||  | ||  ||||| |||  |  ||| |||| |||| | | |||  ||  || ||
Db  31273 AATTGTGGTGAAAAGGTTAAGGATCGATTTCAATACTGGCCCCCATGCTCTGGT 31332

Qy   2547 GCACAGGTGGCCCCAGGCTCTACGTACAAGCCCCTCCAAGAGGGCCAAGATTAACCGTCCCGGCGTC 2606
               ||  ||  | ||||| | | |  ||| || ||||||||| |||| ||| ||   |||| |
Db  31333 GCAGAAAGTTATACAAGGTAATAATATTTTCTCAAAATC---GTTCGTTCAGTTAATCGACCTGTT 31389

Qy   2607 AAGGAGCTCAAAGGGGTCTTCGGCAAAAGGTGTCGAACTGCAGACGAAAGGCGGTGACA 2666
               |||| || ||| || |  || ||| | ||| |||  || ||  || | ||||  ||
Db  31390 AAAGAGACTAAAAGAGTTGTTTCTTAAAGTGGTTTCCTAAAGGACCAAAATTAATTAGA 31449

Qy   2667 ATCGAGGAGGACGAGGAGAAGTACACGTCGCTGGGCTATTTTGATGAGGAGGGGATCAGTGGTGT 2726
               |   || | || |  || || ||   |||  || ||  ||  || ||  ||| |  || |||||
Db  31450 ATTGACTTGGATATAAAGTATATGCGTACTTCTTTTTTTATGATGAAGAACTTAATGAAAAATGGGTGT 31509

Qy   2727 GTCGAAAAGGGTGACTATGAGGTTATCGTGAGCGAACGACGCAGCAGGCGCAG 2772
               ||  ||||| |   |||  | || | ||  || || ||| |  ||    ||
Db  31510 TCTGAAGAGAGGCGAGTACAACGTTCTTGTTCGGAACCAGCAGTGAAAG 31555
```

```
Qy    329 TCCTTCCATCATGGCTGATATTGATGTTGAGGCCAATCTTGAAGAAGCTCACCCTGGCCGA 388
              ||| |||  | |  ||| |  ||||  ||  ||  ||  |||||| ||| || || |||
Db     46 TCCCTTCTGATTTTGCAAAGGCCAACATTGATGAAAATCGTCGAGCAAATTGACGGTCGATGA 105

Qy    389 GAAGGTCGATCTTGCTGGCTGGTATCGACTTCTGCACACAACAAAGGCTCTCCCCAAGCATGG 448
              ||  | || | ||||| ||  | ||  |  ||||  | || ||  || ||  ||||  | |||
Db    106 GGCCATCAGTCTGACCGGCTGGTGTCGGTTTTCTGGCATACGACCATCGGA 165

Qy    449 AGTCCCCTCTCTCGCCGCTTTACAGATAGTGGCCCCAACGGGCGTAAGAGGGACCAAGTTCTTCAA 508
              ||||  |  ||  |  |||| |  ||| |||| |||  ||| ||| ||  |   || |||| |||
Db    166 TGTGCCCAGCTGTAAAAGGTCAGCCAATGGACCAAACGGCATTGCGAGGCAACCACTCTTCAT 225

Qy    509 TGGCGTCCCCTGCGCCGGCCTGCTTCCCTTGCGGCACCTCGCCGGTTCCACATTCAACCAAAC 568
              ||| ||||| ||| | |  |||| || ||||||||| |  |  ||  ||  |||  | ||||
Db    226 GGGGACCCCCTGCAAAAATGTCTACCTTGCGGCACCTCGCCGGTTCCACAACTTGGGGACCCTGA 285

Qy    569 TCTGCTTCGAAGAGAGGCCAGG---TAAGATGAATGGGCAAAGAGGGCCATCGCTTAAGAGTGCGCA 625
              | ||| | ||| ||||| |||   |||||| ||||| |||||||||||||| | |||||||||||
Db    286 AGTCGTCGAGGAGGTCGGGGTTGAAGCTCCTAGCCCCTGAAGCTAAGCTGCGAGCGGCTTC 345

Qy    626 TGTGATCCGCTGGCCCGGACTATCAACATCCAAACGCTCCCCTCGGTCGGTGGACTGGCCTGGA 685
              | |||| |||  | || | ||| |||| |||||  || ||  |  || ||  |  ||| |||||
Db    346 TCTTAGTTCTTTGCTCCTACTTCCAAACATCCAAAGAAAACACCCCGGGAGGCCGAGTTTCGA 405

Qy    686 GTCGATTGGTGTGAGGAATCCGTTTCCTTGGGCGGACCTGCGCGGCCTCTCATCCGCGCGCAT 745
              | ||| | |||| |||| ||| |  ||| ||| | | ||  |||  |  ||| |   ||
Db    406 GAGTTTCTCTGAGGATCCCTACCTTTCGGGTATCATTTCGGCGGGGTCCTACGTCAAATGGTGT 465

Qy    746 TCAGAGCACTGGAGTGCAGGCTCACGATCAAGCACTTTTGTCCAATGGATCAGGAGGACAG 805
              ||||| | |||| || | |||| || ||||| ||||||||||  || |||||| ||||| |
Db    466 GCAGAAGGGTGGAATTGGTGCTACTATCAAACACATTTTGTCGGAAACCACAAGGAGGACGA 525

Qy    806 GCGCATGATTGGTTGCAGAGCATCGTCACGGAGGGAGCGCTCCCGGTGAAATCTACGCACTCCC 865
              || ||||| || | |||| ||| | ||||  ||| |||||||||||  |||||| | |||||||
Db    526 TCGTCAAGGGATAGGATACGGATAGTATTATCACGGAACGTGCTCTCCCGGAAATCTACTCACTCCC 585
```

```
Qy    866 GTTCCAGATTGCTGTGCGAGACTCCCAGCGGGGTGCCGTTCAGTGACGGCCAT  925
          ||||  |                |||||
Db    586 TTTTATGTTAACTCAAAAGTATGCTGCACGCAATCATCGACCGGCTACAACCGAGT  645

Qy    926 CAATGGCGTGTCGTGCGAGCGAGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATG  985
          ||||  ||                                              ||||
Db    646 GAACGGCGTTCATGTAGCGCAAGACCCCTTCCTCGTCAAGCAGGTCCTTACGGTAATGAGTG  705

Qy    986 GGGTTGGGATGGCGTAATCATGACGACTGGGTACGGCACATACAGTACCACACAGAAGCCGT  1045
          ||| |  ||||||||  ||
Db    706 GAAATACAAGGGATTGATCATGACTGACTGGGTTCGGCATGTACAGCGTTGATCATGGGTAT  765

Qy   1046 TGTGGGCAGGCCTCGACCTCGAGATGCGCCGGGACCTCCACGCTTCCGACGGA--GAAACACT  1102
          ||                                                    ||||||||
Db    766 CAAGGCCGGTCTACACCTCGAAATGCCTGGCATCAATAAATGCGTACACTTGATCTAGT  825

Qy   1103 CAAGTTCAACGTGCTCCAACGGAAAGCCCTTTATCCCACGTCAATAAATGGCGTACACTGAGGA  1162
          |||
Db    826 CAACCGTACTATTCAAGCTCGCAAACTCACCCACGAGACATCAAAGGACCGAGCACGAGT  885

Qy   1163 AGTTCTTCAGTTCGTTCAAGAAGTGTTGCT-------GCCTCCGGAGTGACGGAGAACGGCCC  1216
          |||| ||||   |||||||||||||
Db    886 TGTATTAGAGTTAGTCAAGAAGTGTTGCTCAAGGACCCCAGAGAATTCTCGATGGCGATGG  945

Qy   1217 CGAGACGACTGTCAAACACACCCCGAAAACGGCAGCTCTTCCTCCGGGAAGGTTGGCCAACGA  1276
          |                                                  ||||
Db    946 CGAGGAGCGTTACAGTCGCGAGCTTGAATCACCAGCTGCTCATGCGGCGCATCGCATCTGA  1005

Qy   1277 GGGCATACGTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGA-------GCAAGAA  1327
          |||
Db   1006 GTCCAATCGTCTACTCAAAAATGATAACGTTCTCCCTCTTAAACCTGAAGGAGGGCCTC  1065

Qy   1328 GAAGAAGAACGCTGATTGTTGCTGGCCCCAAACGCCAAGCAGCCGCACAATACCACGGGCGAGGAGCTC  1387
          ||||||| |
Db   1066 CAAGAGAATCGGCCGTCGTTGGTCGGCAGGGCGGGAGCTC  1125

Qy   1388 TGCCGCCACTCAGGGCCTACTACGGACGAGTCACTCTCCCTTTTGACGGCGTTCAGCAGCTCGA  1447
          ||||
Db   1126 TGCGGCGGCTGAAAGGCCGGTCATTACTTCAACTCTCACCCTACGAGGGTATTAAGGCCGGCCCTCGA  1185

Qy   1448 GACGCGCCCATCGTTACACC-------GTCGCGGCCTACACCACCGGTTCCTCC  1492
          ||||
Db   1186 ACCTCATGGGGTAGAGGTCACATTCAGTGTGAGGGGTTGCTGCCTATAAAACACTACCCAC  1245

Qy   1493 CATTCTTAGGCGCAGCAGTGCCTCCACCGCCCGACGGCGGCTCCGGGCATGCGGCTGGAGGGCTTT  1552
          ||
Db   1246 CCTCGAATGGGAATCTGGAAAACCGAGACCGCGGGACGTGGCACGTGCCACAC  1305

Qy   1553 CAACGAGCCCCCTGGTACC-------CCTAACACCGGCCACCGACACATTGACGAGCTCTTTCTT  1603
          |   |
Db   1306 CCCAGGAGTCCGACGCACGATGAGCAATGCCATTCATCGCTCCCGATTGGT  1365

Qy   1604 CACCAAGAGACGGACATGCCACCTGGTGGACTACTACCACCCCAAAGGCGGCACACACTGGTA  1663
          |   |
Db   1366 TGACGAAACAGAAATCTTTCATCAGCACGACGTCGTATCCCAAAGGGAATTACAAAGGAGGTGGAC  1425

Qy   1664 CGCCGACACGGGAGGG----CAGTACACCGCCGACGAGGAGCTGCACCTCACCTACGAGCTCGGGCCT  1720
          ||
Db   1426 GATGAGACTGAAGGGCTACCTGAAGCGCACGCTTACCTTCGAATTTGGTGTCT  1485

Qy   1721 CGTCGTCTGCCGCGCACGGCCAAAGGCGTACGTGAGACGACCAGCTCGTCGACAACGCCAC  1780
          |
Db   1486 TATTGCTGCTCGGTCCGGGCTTGACGGAGACAACTGGTTAATCGACAACAACTGGAC  1545

Qy   1781 CAAGCAGAGTGCCCGGGCAATGGCGTCCTTCTTCGGGCTGCCACCGCGAGGAGACGGGCCGGGCCAT  1840
          |
Db   1546 CCGTCAACGTCGTGGTCTGGGCTGAAAGCAGTTCTTTGGGCAGTGGTAGTCAGGGAGAAAACTGGGGTCTA  1605

Qy   1841 CAATCTCGTCAAGGGCAACACGTACACAAGTTCAAGAATCGAGAGTTCGGCCTC-----CGCACC  1894
          |
Db   1606 TCTCCTCAAAGGCTGGCAAGAAGCACATGAAAATCTATGTGTGCGAGTTACTGTAATGTCGCGCGCCTCC  1665

Qy   1895 CACCTACACCCCTTCAAGGGCGACAACCATCGTCCCCGGCCACGGCGTCCTCCTCGCGGGTCGGGCGG  1954
```

```
           Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
           Actinomycetales; Micrococcineae; Cellulomonadaceae; Cellulomonas.
REFERENCE  1  (bases 1 to 3578)
  AUTHORS  Wong,W.K., Ali,A., Chan,W.K., Ho,V. and Lee,N.T.
  TITLE    The cloning, expression and characterization of a cellobiase gene
           encoding a secretory enzyme from Cellulomonas biazotea
  JOURNAL  Gene 207 (1), 79-86 (1998)
  MEDLINE  98172737
  PUBMED   9511746
REFERENCE  2  (bases 1 to 3578)
  AUTHORS  Wong,W.K.R., Ali,A., Chan,R.W.K., Ho,V. and Lee,N.T.K.
  TITLE    Direct Submission
  JOURNAL  Submitted (27-MAY-1997) Biochemistry, The Hong Kong University of
           Science and Technology, Clear Water Bay Road, Kowloon, Hong Kong
FEATURES             Location/Qualifiers
     source          1..3578
                     /organism="Cellulomonas biazotea"
                     /strain="ATCC 486"
                     /db_xref="ATCC:486"
                     /db_xref="taxon:1709"
     gene            327..2813
                     /gene="cba"
     CDS             327..2813
                     /gene="cba"
                     /function="hydrolyzing beta-D-glucoside and cellobiose"
                     /note="one type of cellulase"
                     /codon_start=1
                     /transl_table=11
                     /product="cellobiase"
                     /protein_id="AAC38196.1"
                     /db_xref="GI:2921740"
                     /translation="MTSQTALDPAALVASLPLETKVRLLTGATAFTLAPEESIGLGEV
                     RLSDGTGVRGLKFSGGRTVALFPNATLLASAWSEESTEVGRLLAEEALAQQIHVVL
                     GPTINLHRSVLGRLFEAYSEDPLLTGRLAAAYVRGLQDLGVGACLKHLVANESETER
                     NTMNSVVDPATLRELYLLPEIAVDESDPWSVMAAYNDVNGVPATEHHIVVNEVLKGE
                     WGYGLVMSDWFAATRTRNAPAAPAAGGILDLVMPGPDGPRGDALVAAVRSGELDESVVDDHL
                     RRLIVLAARVGALGDLRDYPDDLPAPDSAVRREQITRLAAAGMTVLTNADDTLPLARG
                     TRVALVGRHALETIDMGGGSATVNPPYQVSVAEGLTALLGDAVDVVDGVEVRTRPVPA
                     RPGFYVVDPDIGRPGLHILTLLAADGTVLDERHDAPSTVWWGFDDDFPQAVARVERARV
                     AGEGALEVGAIGVGRWQVTAGGTELAWTLATSGYGFAEEMLAPTTRTDQVHVGSDAVV
                     DATVVLRSSTRSRSVTVGDADPGTDAGAAAEPLAGVGLFGLVARPAFEAEDVITRAAAA
                     PAQADVAVVVVGLIEEEESVDKSTIALPGADALVRAVAAAARRTVVVVWNAATPVL
                     MPWLDDVDAVLWAGLPGQEEGHAVAAALLGDQEPTGRLVTTFPAADGAAPAWSTVPVD
                     GDLEYTEGRRVGYRGHWADRAPAPAFWLCHGLGYATWEYADAILDTDGDAPAVTVTVT
                     NTGARTSRSRREVQVYVLEPASSDEPVRLVGWADATVDAGASARVTVTADARMWRRWDEAA
                     GGWSRLADGGRLIVARGLGDVRATLALPTA"
BASE COUNT      426 a    1421 c    1306 g    425 t
ORIGIN
```

```
Query Match             11.88;   Score 350.4;  DB 1;   Length 3578;
Best Local Similarity   48.7%;   Pred. No. 1.2e-36;
Matches 1087;   Conservative   0;  Mismatches 1121;   Indels   24;   Gaps   4;

Qy    447  GGAGTCCCCTCTCCTCCGCGTTTACAGATGGCCCCAACGGGGTAAGAGGGGACCAAGTTCTTC   506
           |||    | |||||   | |||||||| |||||||||| ||||||  ||||| |||||||||||
Db    444  GGCCTCGGGCGGAGGTCCGCCTGTCCGACGGCCCGACCAGCGGGCTGCGGGCCGCTCAAGTTCTCC   503

Qy    507  AATGCGTCCCTGCGGCGTCCCTCGCTTCCCTTGCGGCACGTCGGTTCCACATTCAACCAA   566
           ||       |||    ||||| | |||| || || || ||  ||| | |||||  |||||
Db    504  GGCGGGGGCGACCGTCGCGTGTTCCCAACGACGACCTGCTCCCGGTCCCCGTGAGCGAG   563

Qy    567  ACTCTGCTCGGAAGAGGCAGGTAAGATGATGGGCAAAGAGGCCATGCGTAAGAGTGTGCGCGCAT   626
           ||   || || || | ||  ||   ||| | |||  || || || || || || |||||||
Db    564  GAGTCGGACGACCGAGGTCGGCCGCTGCCTCGCCAGAGCAGGCCGCAGAGATCCAC   623

Qy    627  GTGATCCTCGGCCCGGCCGCGACTATCAACATGCAAAACGCTCCCCTCTCGGTGGACGTGCGCTTCCAG   686
           ||||| || ||   || || ||||| | ||| || ||  ||| ||| |||    ||||| || |
Db    624  GTCGTGCTCTCGGCCCCCACGATCAACCTGCACCGGGTTCGGCCGCTC   683

Qy    687  TCGATTGGGTGAGGATCCGGTTCCTGGGCGCGGCTTGGGAGCTGGCGGCGCTCATCCCGGCGATT   746
           |  | ||        |||   |||| ||||||||||| |||||||||
Db    684  GCCTACTCCGAGGACCGGCGTGCTCACCGGCGCGCCGGCCCCCTACGGTGCCCGCCTG   743

Qy    747  CAGAGCACTGGAGTGCAGGCTACGATCAAGGAGGAGACAG   806
           |||  ||    | |||  ||||||  |||||||| || |||
Db    744  CAGGACCCCTCGGGCGGGTCGGGCGGCCTTGCCTCAGCACCACTTTTTGTGCAAGATCAGGAGGAGACAG   803

Qy    807  CGCATGATGGTCGCAGACGCATCGTCACGGACGGGGGCTCTCCGTGCGACCAACGAGTCGGAGACCGAG   866
           |||||  |||||||| ||||   || | || || ||||   ||| | ||||| ||    ||||  |
Db    864  TTCGAGATCGGCGTCGACGAGTCGAAAGTGGTGCCGTTCATGACGCGCGTACAACGAGGTC   923

Qy    867  TTCCAGATTGCTGTGCGAGACTCCCAGCGCGGGGGCGGTGCGTTCATGACGCGGCGTACAATGGCATC   926
           |||||||||| || || || || || || |||||||| ||||||||||||| ||||  || |||||
Db    924  AACGGGCGTCCCCGACCAGGACCACCACCACGTCGTCAACGAGGTCCTCAAGGGCGAGTGG   983

Qy    927  AATGGCGTGTCGTTGCAGCGGAGAAACCCTAAAATATCTTGATGGGATCGTTCGAAAAGAATGG   986
           |||||||| |||| ||| ||| || |||||  ||| ||| ||||||||| | |||| | |||
Db    984  GGCTACACCGGCGCTCGTCGAGTGCCAGACTGGTTCGCGGACCGGCGCCGCCTG   1043

Qy    987  GGTTGGGGATGGCCTAATTCATGACGGACACTGGTACGGCGACATACAGTACCACGAGAAGCCGGTT   1046
           || |||||||||||||  || || || ||||| |||||||||| | ||||||||||| |||||
Db    1044 GCCGCGGGGGCTTCAGACCTCGTCATGCCCGGCGACGCCCCTCGGGGCGGGCGGCCGGGCCTCGTC   1103

Qy    1047 GTGGCAGGCCTTCGACCTCGAGAATGCCCGGACCTCCACGCGTTCCGGAGGAGAAACACTCAAG   1106
           || ||| || || |||||||  |  ||||| || || ||  ||  ||||| | ||| || |||
Db    1104 GCCGCGGCGTGCCGCAGCCGCGAGCTCGACGAGTCCGGCGTCGTCGACGACCACCTGCGCCGCCTG   1163

Qy    1107 TTCAACGTCTCCAACGGGAAAAGCCCTTTTATCCACGTGCATTGACCCAGAGGGGCT:AGGGAAGTT   1166
           |||||| |  ||||  | ||||||||||| || || || || || | |||||| | ||||||
Db    1164 CTCGTCTCCGGCCACGCGTGCTGCAAGAAGTGTGCTCCGCCCTCCGGGAG---TGACGGAGAACGGCC:CCGAGACG   1223

Qy    1167 CTTCAGTTCGTTCAAGAGAGTGTGCTCCGCCCTCCGGGAG   1223
           ||||||||||| ||||||||||||||||||||||||||||
Db    1224 ACTGTCAACAACACACCCGAAAACGGCAGCTCTCCTCCGGGAAGGTTGGCAACGAGGGGCATC   1283

Qy    1224 ACTGTCAACAACACACCCGAAAACGGCAGCTCTCCTCCGGGAGGTTGGCAACGAGGGGCATC   1283
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1284 CCCGCCCCCGACAGCGGCCGTCCGGCGGCGAGCTCACCGGCGGCCATCACCACCGGCAAGATCAAGAGAG   1343

Qy    1284 GTGCTGCTGAAGAAGGGAGGAGCCACAATACCACCGGCAAGATCAAGAGAGAGAAGACCCTGATT   1343
           |||||| || |||||||  || |||| ||||||||||  |||||||||||||||||||||
Db    1344 ACCGGTCCTCACCAAGCCGACGAACACCCTGCCGCTGCGCGGGGCCCTCGGCGGACCACTC   1403

Qy    1344 GTCGGCCCCAAACGCCAAGCAGGCCCACATACCACCGGCAAGAGCCTCTGCCGCCACTCAGGGGC   1403
           ||||| |||| | ||||  | ||||||| || || || || || |||  |||| | |||||
Db    1404 GTCGGGGCGGCGCACGCGCCTCGAGACCAATCGACATGGGCGGGCTCCGGCGACGGTCAACCCG   1463

Qy    1404 TACTACGCAGTCACTCCCTTTGACGGGCGCTCAGCAAGCAGCAGTCGGCGGGGGGCCCATCGTAC   1463
           |||  | |  | ||||  |||||| || |||  || || || ||   ||||||   ||||||
Db    1464 CCGTACCAGGTGAGCGTTCCGCGAGGGCCTCACCGGCGCTGTGACGTC   1523

Qy    1464 ACCGTCGGGCGCCTACAACCACCGTTCCTCCCATTCTAGGCGGAGCAGTGCCTCACGCCCGAC   1523
           ||||||||  ||  ||| | |||   ||||||| |  ||| ||||| || ||||||||||||
Db    1524 GTCGACAGGCCGGTCGAGGTCCGCACCCGCCCCGCGGGCCCGCCGGGTTCGTTCGGGAC   1583

Qy    1524 GGCGCTCCGGGCATCGGCTGAGGGGTCTTCAACGAGGTCCCCGCGGGCCCTGGTACCCCTAACGGCAG   1583
           |||||||| |||||| ||||| |||  || ||||| ||  |||||||||| || ||   | |
Db    1584 CACATTGACGAGCTCTTCTTCACCAAGAGCAGCAGACATGCACCTGGTTGGACTTACTACCACACCC   1643

Qy    1584 CACATTGACGAGCTCTTCTTCACCAAGAGCAGCAGACATGCACCTGGTTGGACTTACTACCACACCC   1643
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1644 GACGAGGCGGCGACGAGGCCGTCGACGACACGGCGATGTGGCGGCGTTGGGACGAGGCAGCGGCAGTGC   1703

Qy    1644 AAGGCGGCAGAACACGTGGTTACGGCGCGACATGGAGGGCAGCGTACACCGGCGCGGCCCGAGTGC   1703
           |||||||| |||||||  | ||||||| | || |||  |||| || || || ||||| ||
Db    1704 GCGATCGGCGGGTCGGGCCGGCAGTTGACCGGCAGGGCGACCGGCACCGGCGGGCTC   1763

Qy    1704 ACCTACGAGGCTCGGCC--------TCGTCGTCTGCGGCGACGGCACCGGCAAAGGCGTACGTAGAC   1754
           ||||||||||| |||||        ||||| || || |||||| ||| ||    |||  ||||
Db    1764 GCGACGTCCGGCCACCGGGGTTCGCCGGAGGATGCCTCCGCGGAGCTCGCCGGTCTTCGGCGTC   1823

Qy    1755 GACCAGCGTCGTCGTACGAAACCGCCACCAAGCAGGGTCGGCGGTGCCTTCTTCGGGTCC   1814
           |||| || |     |  ||| || || ||   || |  ||  ||| || |||||
Db    1824

Qy    1815 GCCACCCGCCGAG---GAGACGGGCGCCGCGATCAATCTCGTCAAGGGCAACAC3TACAAGTTC   1871
           |||||||||||||   || | |||||||||||||  ||  ||||||| ||||  |||| |||
Db    1871
```

```
RESULT 8

LOCUS       ATUCBG1      3710 bp    DNA    linear   BCT 26-APR-1993
DEFINITION  A.tumefaciens beta-D-glucosidase (cbg-1) gene, complete cds.
ACCESSION   M59852
VERSION     M59852.1  GI:142221
KEYWORDS    beta-D-glucosidase.
SOURCE      Agrobacterium tumefaciens
  ORGANISM  Agrobacterium tumefaciens
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 3710)
  AUTHORS   Castle,L.A., Smith,K.D. and Morris,R.O.
  TITLE     Cloning and sequencing of an Agrobacterium tumefaciens
            beta-glucosidase gene involved in modifying a vir-inducing plant
            signal molecule
  JOURNAL   J. Bacteriol. 174 (5), 1478-1486 (1992)
  MEDLINE   92165721
   PUBMED   1537792
FEATURES             Location/Qualifiers
     source          1..3710
                     /organism="Agrobacterium tumefaciens"
                     /db_xref="taxon:358"
                     /cell_line="B3/73"
                     /tissue_lib="genomic HindIII cosmid"
     gene            169..2637
                     /gene="cbg1"
     RBS             169..173
                     /gene="cbg1"
     CDS             181..2637
                     /gene="cbg1"
                     /EC_number="3.2.1.21"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-D-glucosidase"
                     /protein_id="AAA22082.1"
                     /db_xref="GI:142221"
                     /translation="MIDILDKMTLEEQVSLLSGADFWTVAIERLGVPKIKVTDGPN
                     GARGGSLVGGVKSACFPVAIALGATWNDPELIERAGVALGGQAKSKGASVLLAPTVNI
                     HRSGLNGRNFECYSEDPALTAACAVAIYINGVQSQGVAATIHKFVANESEIERQTMSSD
                     VDERTLRIYILPFPEEAVKACVKKAGVKAVMSSYINKLNGTYVSENPWLLTEKVREEWGPGCV
                     VMSDWFGGSHSTAETINAGIDLEMPGPWRDRGEKLVAAVREGKVKAETVRASARRILLL
                     LERVGAFEKAPDLAEHALDLPEDRALIRQLGAEGAVLLKNDGVLPLAKSSEDQIAVIG
                     PNAASARVMGGGSARIAAHYTVSPLEGIIRAALSNANSLRHAVGCNNNRLIDVFSGEMT
                     VEYFKGRGPESRPVHEIEVEKGEFFWFDLPSGDLDLADFSARMTATFVPQETGEHIFG
                     MTNAGLARLFVDGELVVDGYDGWTKGENFFGTPANSEQRRAVTLGGARRYRVVVEYEAP
                     KASLDGINICALRFGVEKPLGDAGIAEAVETARKSDIVILLVGEGEWDTEGIDLDPDM
                     RLPGRGEELIEAVAETNPNVVVVLQTGDPIEMPWLGKVRAVLQMWYPGEGLGNALADV
                     LFGDVEPAGRLPQTFPKALTDNSAITDDPSIYPGDQDHVRYAEGIEVGYRIHDTREIE
                     PLFPFGFGLGYTRFTWGAPQLSGTEMGADGLTIVTDVDTINIGDRAGSDVVQLYVHSPNA
                     RVEBPFKELRAFAKLKLAPGATVIAPRDLAYFDVEAGRFRADAGKYELIVAAS
                     AIDIRASYSIHLPVDHVMEP"
     misc_feature    213..223
                     /gene="cbg1"
                     /note="active site"
BASE COUNT      718 a   1084 c   1168 g    740 t
ORIGIN

Query Match     11.5%;  Score 343.4;  DB 1;  Length 3710;
Best Local Similarity   49.1%;
Matches 1195;  Conservative   0;  Mismatches 1171;  Indels  69;  Gaps  8;
```

```
Qy      355 TTGAGGCCATCTTGAAGAAGTCACCCTGGCCGAGAAGGTCGGATCTGCTGGCGTGGTATCG 414
                |      |         |  |    |   |  | | |  |      |  |   | ||  |
Db      185 TCGACGATATTCTCGATAAGATGACACTGGAGGAGCAGGTGTGCCCGTGGCTGGCGCGG 244

Qy      415 ACTTCTGGCACACAAAGGCTGTCTCCCCAACATGGAGTCCCCTGTCTCCGCGTTTACAGATG 474
              | |||  |             |       |     |        |   |   || | |
Db      245 ATTTCTTGGACGACACCGTCGGATCGAGCGGGCTGCCGAAGATCAAGGTTACCGAC 304

Qy      475 GCCCCCAACGGCGTAAGAGGGGACCAAGT----TCTTCAATGGCGTGCCCTGCGGCCTTCC 531
             |||||   |  ||| ||  |           |  |   |       || |||||   |
Db      305 GCCCCCAAATGCGCACGCGCGGGTCGGCTGGGTCGGCGCCCTGCAAGTCGCGCTTGCTTCC 364

Qy      532 CTTGCGGCACCTCGGCTTCCACATTCAAACCAAAACTCTGCTCGAAGAGGCAGGTAAGA 591
             |  ||  || |  ||  |   |      |   |  | | | |    |  |  ||
Db      365 CGGTGGCAATCGCGCGTTGGAGCGACGTGGGACCCGGAGCCTCATCGACGCGGCCGCGTGG 424

Qy      592 TGATGGGCAAAGAGCCCATCGCTAAGAGTCCGCATGTGATCCTGGCCCGACTATCAACA 651
             | ||    |   |        |      || |   |  || | |  |    |  || |
Db      425 CGCTCGGGAGGACAAGCCAAGACAAGGGCGCGGCTGCGTCGGGTGTCTTCGGGCGCGTCAACA 484

Qy      652 TGCAACGCTCCCCTCTCCGGTGGACGTGGGCTTCGAGTCGATTGGTGGACAATCGGTTCCTGG 711
             | |    |  ||  |  ||  || ||  | |   || |     |  ||   || |   | ||
Db      485 TTCACCGCTCCGGCCTCAATGGCCGCGCAACTTCGAATGCTATTCGGAAGAGACCGCGCCGCTGA 544
```

Alignment (genomic region, Db/Qy blocks):

```
Db   1824 GTCCACGTGGGGTCCGACGGGTTCGTCGACGGCGTGCTGGCGTCTGCCACCGC 1883
Qy   1872 AAGATCGAGTTCGGCCTCCGCACCGCCACACCTCCAAGGGCGACCATCGTTCCCCGGC 1931
Db   1884 AGGGTGACCGTCGGGACGGCCGCCGCCGGCTCCCGC ... 1943
Qy   1932 CACGGC------TCCCTCCCGGTCGTCGGCTGGCTCAAGGTCGCTCCGTCC 1982
Db   1944 GCGGGGTTCGGGCGTGTTCGGGCTTCTCGGCGGCG ... 2003
Qy   1983 ATGGAAAAGTCCGTCGCCCTCGCCCAAGAGACACGACAGCGCC 2042
Db   2004 ATCACCCCGCGCCCCGGCGGCGGCGGCGGCGGC ... 2063
Qy   2043 AAGGCCGACTGGGAGACCGGAGGGCGGCGACCTGC ... 2102
Db   2064 ACCGAGGAGGAGGAGGACGAGGAGTCCGTCGACAAGTC ... 2123
Qy   2103 GACCAGCTCATTGCCGACGTGGGGCGCCGGCCGGAG ... 2162
Db   2124 GACGCGCTCGTGCGTGCGGGCGTCCCCGGGCCGC ... 2183
Qy   2163 GGCACCCCGGAGGAGATGCCCTGGCTCGACGCGCCCACGC ... 2222
Db   2184 CCGACGCGCCGGTGCATGCGCTGGCTCGACGACGTC ... 2243
Qy   2223 GGGGGCAAGCAGACGAGGGCAAACTCCATTCGCGCGACGTCTTC ... 2282
Db   2244 CCCGGCCAAGGAGGGGGCGCGGGGCCGTGGCCGGGC ... 2303
Qy   2283 GGCAAGCTGTCCCTTCCCAAGCGGCACGACAACCCCGGCG ... 2342
Db   2304 GGCGGCCTCGTCACGACGTTCCCGCGCGGCGAGGACGC ... 2363
Qy   2343 CGCACCGAGGCCGCGGGGCGCGAGGCGCACGCGTCGTACGG ... 2402
Db   2364 CCCGTCGACGCGGGACCTGGAGTACACCGAGGGCGGCCACTGG 2423
Qy   2403 GAGTTTGCCGACAAGGACGCTCAATTTCCCCTTTGGCCACGGCCTG ... 2462
Db   2424 GCCGACGGGCGCGGCCTTCGGGCCCGGCGCCTTCGGGTTACGCGACGTGG 2483
Qy   2463 GCCTTTTCCAAATCTCTCCGTGTCTCACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTG 2522
Db   2484 GAGTAGGCCCACGGCGTCGACACGGCACGGCAAGCTACGCCGTCACCGTG 2543
Qy   2523 AAGAGACACCGGCTCCGTGCCCGGGCCACAGGTGCAGCACCCCTCCAA 2582
Db   2544 ACCAACACCGGCGTTAACCGCCCGGAGGGTCTTCGCCAAAGGTCGAAACTGCAG 2603
Qy   2583 GCGGCCCAAGATTAACCGCCCCGTCAAAGGAGCTCAAGGGGTTCGCAAAGGTCGAAACTGCAG 2642
Db   2604 TCCGACGAGCCCGTGCGCCTTCGTCGGGCGACGGCAGAACCGGGCGTCC 2663
Qy   2643 CCCGGCGGAGACG 2654
Db   2664 GCCCCGCGTCACG 2675
```

```
Qy   712  CGGGCTTGGGAGCTGCGGCGCTTCATCCGGCGCATTCAGAGCACTGAGTGCAGGGTACGA  771
Db   545  CGCGCGGCCTGCGCCGGTGCGCTATATCAATAGGCGGTGCGAGACCATCGCCGCACGA  604

Qy   772  TCAAGCACTTTTGTGCAATGATCAGGAGGACAGGCGGCCATGATGGTGCAGAGCCATCGTCA  831
Db   605  TCAAGCACTTCGTTCGCAACGAGTCGGAGATCGAGCGAGCCACCATGTCCTCCGATGTCG  664

Qy   832  CGGAGCGGGGCTCTCCGTTGAAAATCTCAGCACTCCCCGGTTCCAGATTGCTGTCGCGAGACTCCC  891
Db   665  ATGAGCGGACGCGTCGCCGAAAATCTATCTGCCCGCCTTTCGACGGAGGCGGGTGAAGAAGGCCG  724

Qy   892  AGGCCGGGTTGCGTTCATGAGACGGCGTACAATGGGTGTCGTGTGCAGCGAGAACC  951
Db   725  GCGGTGAAGGCCGTCAATGTCCTCCTACAACAAAGCTCAACGGCAACCTATACGGCCG  784

Qy   952  CTAAAATATCTTGATGGGATGGCTTCGAAAAGGAAGTTGGTGGATGGCCTAATCATGACG  1011
Db   785  CCTGGCGTTGACGAAAAGTCCTGCGGGAATGGGGCTTCGACGGCGTGGTTCATGTCCG  844

Qy  1012  ACTGGGTACGGGCGACANTACCACAGANACCCGTTGTGGCAGGCCTCGACCTTCGAGATGC  1071
Db   845  ACTGGTTCGGCGGCTCGCACTCGACGGCTGAAAACCATCAANATGCCGGGTTCGATCTGG  904

Qy  1072  CCGGACCTTCCACGCTTCCGAGGGAAACACTCAAGTTCAACGGGAAAGCCCT  1131
Db   905  CGGGGCCCTTGGCGGGATCGCGGGCGAAAAGCTTGGCGGGCAGCTGCGGGGAAGGCAAAGGTAA  964

Qy  1132  TTATCCACGTCATTGACCAGGAGGGGCTAGGGAAGTTCTTCACTTTCAGTTCGTCAAGAAGTGTGCTG  1191
Db   965  AGGCCGAGACGGTGCGCGCTTCGGCGACGGGGTATTCTCCTTCGTCTGCTCGAACGGCGTCGGCG  1024

Qy  1192  CCTCCGGAGTGACGGACATACAGTACCACAGANACCGTTGTGGCAGGCCTCGACCTTCGAGATGC  1251
Db  1025  CCTTTGAAAAGGCGCCTGAACGCGGCGTGAAAACCATCAAATGCCGGGTTCGATCTGGACATGC  1084

Qy  1252  CTCTCCTCCCGGAAGGTTTGGCAACGCATTCGTCGTGCTGAGACGAGAACAACGTTC  1311
Db  1085  TCATCCGCCAACTCGGTGCCGGAGGGTGCTTACTCCTGAGAAATGACGGAGTGCTGCCGCG  1144

Qy  1312  TGCCCTTGAGCAAGAAGAAGAGAGACGGCTAGGGAGAAGTTCGGCGGGCCCCACGCGCAGCGGCCACAT  1371
Db  1145  TTGCCAAGTCGTCCTTCGACCAGATCGGCCCTCATCGGCCGCTTGCCGCACGCG  1204

Qy  1372  ACCACGGCGAGAGGCTCTGCCGCGGCCTACTCACTCCCTTTGACGGCC  1431
Db  1205  TCATGGGCGGGAGGAAGCGGCGGGATTGCGCGCATTTATACGGTGAGCCCCACTTGAGGCCA  1264

Qy  1432  TCAGCAAGCACGCTCGAGACCCGCCATCGTACACCGTCGCGGCGCCTACACCACCGGTTCCTC  1491
Db  1265  TTCGCGCGGGGCCT- - - - - -GTCCAACGCAACAGCCTCCCCATG  1303

Qy  1492  CCATTCTAGGCGGACCAGTGCCTCACGCCCGACGCGCCGTCCGGGCATGCGCGCTGGAGGGGTCT  1551
Db  1304  CGGTCGGGCTGTAACAAACAACCGGCGTCATCGACGTCTTCAGCGGGATGATGAGACGGTGGAAT  1363

Qy  1552  TCAACGAGCCCCCTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTCTTCTTCACCAAGA  1611
Db  1364  ACTTCAAAGGGACGCGGCGGCTTCGTCCANGTCGAGACGCGTACGCGCGACA  1423

Qy  1612  CGGACANTGCACCTGGTTGACTACTACCACCCCAAGGCGGCAGACGTGGTACGCCGACA  1671
Db  1424  AATTCTTCTGGTTCGACTACTTCCGTCGCGGCGGATTTTTCGGCGCCA  1483

Qy  1672  TGGAGGGCACGTACACACCGCGACGAGGACTGCACCTACGAGCTCGGCCTCGTCGTCTGGG  1731
Db  1484  TGACGGCGCAAAGCGGCGACCTTCGTGCGCGACGAGAAACCGGTGAACACAATCTTCGGGCAANGGTTGGACA  1543

Qy  1732  GCACGGCAAAGGCGTACGGTAGACGACCAGCTCGTCGTCGACACGCCACCAAGCAGGTCC  1791
Db  1544  GGCTTGCTCCGGCGTGTCGTGGACGGCGAACTGGTTGGTCGGACGGCTATGACGGTTGGACGA  1603

Qy  1792  CCGGCGGATGCCTTCTTCGGCTCCGCCACCCGCGAGGAGACGACGGGCGCATCAAATCTCGTCA  1851
```

```
Db  1604  AGGGTGAGAACTTTTTGGAACGCGGCAANCACGGCGAACAGCAGAGGCGGTAACGCTTGGGG  1663

Qy  1852  AGGGCAANCACGTTACAAGTTCAAGATGCGGAGTTCGGCGTTCGGCGACCACCTACACCCTCAAGG  1911
Db  1664  CCGGCACCGCGTACCGGTGTTGGTGTCGGAATATG- - - - - -AGGCGCCGGAAGG  1708

Qy  1912  GCGACACCATCGGTCCCCGGCCACGCGTCCCTCCTTCCCGGTTGGCCAAGGTCATTGACG  1971
Db  1709  CCAGCCTGGACGGGCATCAACATATGTGGCGGTCCGGGCTTCGTGTGTCAAAAGCCCGT- - -CG  1765

Qy  1972  ACCAGGCGGAAATCAGAAAAGTCCGTGCCCTGCCAAGGAGCAACGGAACGACCAGGTCATCATCT  2031
Db  1766  GCGATCGGGATTGCGGGGTCGAAAACGGTCGGCAAGTCCGATATCGTACTGCTCC  1825

Qy  2032  GCGCGGGGCCCTTAACCGCCGACTGGGAGACGGGGCGACCCGAGGCAGCATGAAGCTCC  2091
Db  1826  TCGTCGGGGCCGTGAGCGGCCAGGAACTTGGCAATGCGGACACGGGGTCGCGGGGGCTTGGCGGACT  1885

Qy  2092  CCGGCGTGCTGGACCAGCTCATTGCCGACTGGGCGGGGGCCCCGGGAACCCAAACACCCGTCGTCG  2151
Db  1886  CGGGTCGCGCAGGAGGAGCTGATCGAGGCGCGAAAACCCAACGTGTGGTCGTGG  1945

Qy  2152  TCATGCAGACGCGGGGCACCCCCGAGGAGAATGGCCCCGAGGAGGAGTGTACGGCGGAGGACG  2211
Db  1946  TACTGCAAAACGGGTGGTCCATCGAGAATGCCATGGCTCGGCAAGTGCCGCTGTGGGCTGC  2005

Qy  2212  AGGCCTGGTACAGGCGGCCAACGAGCACGGGGGCGAAACTCCATTGCCGGACGTCGTGTCTTTGGCGGACT  2271
Db  2006  AGATGTGGTATCCCGGCCAGGAACTTGGCAATGCGGACAGTTCTCTTTGTGGGCT  2065

Qy  2272  ACAAACCCCTCGGGCAAGCTGTCTCCCTCAGCTTCCCCAAGGCGCTTGGCGGGAACAACCCCGGCT  2331
Db  2066  TCGAGCCGTGCGCTTGCCCACAGAACCTTCCCGAAAGCGCCTCACGGATNAATTCGGCCA  2125

Qy  2332  TTTCTCAAACTTCCGCAC- - - - - - - - - CGAGGGCGCGGGGCCACGCGTGTACGGGCGAGGACG  2379
Db  2126  TTACCGACGATCCGTGATCTATCCTGGGCCAGGAAGTGCGCTACCCGGAAGGGA  2185

Qy  2380  TCTACGTCGGGTGGGTTACACAGTTACTACGAGTTTGCCGACAAGGAGCGTCAAATTCCCCGTTTGGGC  2439
Db  2186  TCTTCGGTCCGTGCTATCGCCATCACGATACAAGAGAGATCGAACACTCTTCCCCTTCGGGCT  2245

Qy  2440  ACGGCCTGTCCTACACCACCACTTTTGCCTTTTTCCAANTCTCTCCGTGTCTCACAAGA- - - - -  2494
Db  2246  TCGGGTTCTTGGCTACACCCGCCTTTTACCTGGGGGTGCCCCGGCTGGAACGGAAATGG  2305

Qy  2495  - - - -CGGCAAGCTCAGCGGTGTCCCCTCTCCGTGAAGAAACACCGGCTCCGTGGCCCGGGCCAC  2550
Db  2306  GGGCGGATGGTCTTACGGTGACGGCTCAGATGTCACCAAATATAGGCGGACAGGGGCGGGGATCGG  2365

Qy  2551  AGGTGGGCCCAGCTCTACGTCAAGGCCCTCAAGGATTAACCGCCCGTCAAGG  2610
Db  2366  ACGTGGGTGCAGCGCTCTATGTTCACGTCTCCCAA- - -TGCCAGGGTCGACGGCGGCCG1TCAAGG  2422

Qy  2611  AGCTCAAGGGCTTCGCAAGAGTCGAACTGCAGCCCGGCCCGGCCGCGGGGCCGAGGGCGACAATCG  2670
Db  2423  AGCTGCGTCGCCTTTGCGAAGCTCAAGCTGGCCCCGGGGGAAGGCGGTGACGGGAT- - -CGGTGC  2479

Qy  2671  AGGGAGCAGGAGAGCTACGTCGCGTGCCGTATTTTTGATGAGGAGGAGGCGGGATCAGTGGTGTGTTCG  2730
Db  2480  TGAAGAATGCCGTCCTCCGCGGACTTGGCTTACTTCGATGTGCGAGGCGGCGGGCTG  2539

Qy  2731  AAAAAGGGTGACTATGAGGTTATCCGTGAGCGGACAGC  2765
Db  2540  ATGCGGGGCAAGTACGAGCTTGATCGTGGGGGCCAGC  2574
```

```
ACCESSION   AJ293760
VERSION     AJ293760.1  GI:9796011
KEYWORDS    beta glucosidase; bgl gene.
SOURCE      Agaricus bisporus
  ORGANISM  Agaricus bisporus
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Agaricaceae; Agaricus.
REFERENCE   1  (bases 1 to 2636)
  AUTHORS   Morales-Almora,P. and Thurston,C.F.
  TITLE     Molecular analysis of the cellulolytic genes in Agaricus bisporus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2636)
  AUTHORS   Morales-Almora,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2000) Morales-Almora P., Microbiology, King's
            College London, 150 Stamford Street, London, SE1 8WA, UNITED
            KINGDOM
FEATURES             Location/Qualifiers
     source          1..2636
                     /organism="Agaricus bisporus"
                     /strain="D649"
                     /db_xref="taxon:5341"
                     /clone="PMRT1615"
                     /country="United Kingdom"
     gene            25..2610
                     /gene="bgl"
     CDS             25..2610
                     /gene="bgl"
                     /EC_number="3.2.1.21"
                     /function="putative hydrolysis of beta-D-glucose"
                     /codon_start=1
                     /product="putative beta glucosidase"
                     /protein_id="CAC03462.1"
                     /db_xref="GI:9796012"
                     /translation="MPPSDFANADIDKIVDALTTDEAILLSAGVGFWHTHAIERLQIP
                     AVKVSDGPNGIRGNHFFMGTPAKCLPSSTAMGATFDRDLLEEVGLKLLAREAKLRSAS
                     VILAPTCNIQRNPLGGRSFSESFSEDPFLSGMLCSAYINGIOKGIGATIKHFVGNDKE
                     DDRTGYDSIHYRTSSGDLSPSVHAGSRNMPLHGAIMTAYNRVNGIHVCENPELLQKI
                     LRDEWKSDVNLMSDWFGTYSVDVGLNAGLDLEMPGLINKWRSLESVNRSIQSRKVTAKK
                     VKERARKVELEVKCAQGAPEILLDGDGLERTLLDSDEDNALMRKPAAAESIVLLKNDKKV
                     LPLDRKSLKKVAIVGGNAKAAVLSGGGSAALKPSYVARPAGLVKALKEVNPNIEVTY
                     SEGARAAKTLPTLDLFTESGQRGWTGSWHAHENDDSLVALPTPIKTQVVDETRMFI
                     SSSBPAGIIRKKWTLRMKGYLKPRRDCKPEFGLPYAAGRAKLFVDGKLVIDNWTRQRRG
                     VEFPCCGSEEETGVVELKAGVKHEIYVDPCNVRGPADGDETETIMDSNPGIRLGGAEV
                     EDPDELLNSAVSIAKDADAVIIVVLGNGDWETEGNDRTTLALPGRTDELVQKVVAANP
                     RAVVTQAGSAITMPWVDQVSSLLHTWYLGNATGDAIADVIFGHNPNPSGKLSLIFPKR
                     LEDVPAHGYFHSEKGKVTAEGLIYGYKHYQHRNIEPELFAFGHGLSYTSFNISDLRLS
                     QPIVSCDQFDLKATVLTNTGNIITGSQVVLYIGLPKTSELTHPWQLRGFKKMRDVK
                     PGESEREVLVMDRLSVSYMDKEWVVENGAVDVRVAFTSEEGVGEGQELLGRFNVEKGF
                     GWRGL"
BASE COUNT   657 a   603 c   682 g   694 t
ORIGIN
```

Query Match   10.3%;   Score 305.2; DB 8; Length 2636;
Best Local Similarity   49.9%;   Pred. No. 9.8e-31;
Matches 1106; Conservative   0; Mismatches 1053; Indels 59; Gaps 11;

```
Db  1349 ATGTTGACGAGACAAGAAATNGTTCATCACGACAGTTCTTCTCCTGTCCGGTTATAACCAAGAAGT
         |||                ||                       |||       |||         |||
Qy  1658 GTGGTACGGCCGACGATGGAGGCGTACACCCGCCGACGAG-GACTGCACCTACGAGCTCG
         |||  |||||||   |||||    ||                |||||||||   |||
Db  1409 GGACGCTTCGTATGAAAGGTTATTGAAAACGGCGAGAGCCTGAGTGTGAATTCGAATTTG 1468

Qy  1717 GCCTCGTGTCGTCCGGCACGGCAAAGCGTAGGTAGAGACGACCAGCGTGTCGTCGACAACG 1776
         |||    ||||||||   |||||||||||||||   |||||||||||||||||||||||
Db  1469 GATTGACGCGCTTCGTGGAAGAGCCAAGCGTGTTCGTCGATNGGAAAACTGGTGATTGACAAACT 1528

Qy  1777 CCACCAAGCACAGGTCCCCGGCCATGCCTTCTTCTTCTTCGCTCTCCGCGACACCGACCGTGAGA 1836
         ||||||       ||||||||||||||||||||||  |||||||||
Db  1529 GGACGCGTCAAACGTCGCGGGTGTTGAGAAATTCTTTGGATGTGGTTCAGAAGAGAGACTGGCG 1588

Qy  1837 GCATCAATATCTCGTCAAGGGCAACACGTACAGAGTTCAAGATCGAGTTC------GGCTCCG 1890
         |||                 |||||||||                        |||
Db  1589 TTGTTGAGTTGAAAAGCGCGGTGTCAACCATGAGAATTTATGTCGACTTCTGTGAATGTTCGTG 1648

Qy  1891 CACCCACCTACACCCTCAAGGGCGACACCATCGGAACGTACAGTTCAAGATCGAGTTC 1950
         |||||||       |||||||||||||||||||
Db  1649 GGCCAGCTGATGGTGATGAAACCGAAACAATCATGGATAGTAACCCTGGTATTCGGCCTTG 1708

Qy  1951 GCGGCTGCAAGGTGCATTGACGACCCGAAATGAAAAAGTCCGTTCGCGCCCGTCCCAAGG 2010
         |||||       |||||||||||||||||||||  |||||||||
Db  1709 GGGGCGCCTGAAGTTGAAGATCCGTATCAAATTCGCGTGTCTCAATCGACAAAAG 1768

Qy  2011 AGCACGACCAGGTCATCATCTMGCGCGGGGGGCCTTAACGCCGCACCGCTCCGGTCG 2070
         |||||||||||||  |||||||||  |||||||||||
Db  1769 ACGCTGATGCCGTTATTATGTCGTTGTCGGTTTAAACGGTGATTGGGAGACAGAGGGTAATG 1828

Qy  2071 ACCGGCCGACGATGAGAGCTCCCCGCCGTGCTGCTGGACCAGTCATCGCTGTGGCGCCG 2130
         |||||       |||||||||||||||  |||||||||||
Db  1829 ACGTACCACCCTTGCCCTTCCGGCCGTACCAGGCGGAATTGGTACAAAAAGTCGTTGCTG 1888

Qy  2131 CGAACCCAAACACCGTCTGGTCGTCATGCAGAGACGGGGGCCCTGGCTCG 2190
         |||||||||         ||||||||  ||||||||  ||||||  ||||
Db  1889 CAAATCTCGTGCTGTTCCCTCTCTACAATACCTGGTATCTAGGTAATGCTAACTATCGCCATTG 1948

Qy  2191 ACGGCACGCGCCGTCGGTTACGGCGGCCAACGAGACGGGCAACTCCATTG 2250
         |||||||||  ||||||||||||||    |||||||||||  ||||||||
Db  1949 ACCAAGTTTCTTCCCCTCTACAATACCTGGTATCTAGGTAATGCTAACTATCGCCATTG 2008

Qy  2251 CCGAGCTCGTCTTTGGGACTACACACCCCTCTCGGGGCAAGCTGTCCCTCAGCTTCCCCAGC 2310
         ||||||||   ||||||   ||||||||||||||||    |||||||||||||
Db  2009 CCGACGCTCATTTTGGGAATCATAAATCCTTCTGGAAAGCTTTCATTGACCTTCCGGAAGC 2068

Qy  2311 GCCTGCGAGGACAACCCCGCGGTTTCTCAACTTCCGGCAACGAGGCGCCGGGCCCACGGTGTACG 2370
         ||||||         |||||||||||||||     |||||||||  ||||||  ||||
Db  2069 GATTGGAGNGANGTTCCTGCTCATGGCTATTTCCATTCGAGAAAGGAAAGGTGACGTACG 2128

Qy  2371 GCGAGGACAGTCTACGGTCGGGGTACAGGTACTACGAGTTTGGCGACAAGGACGTCAATTTCC 2430
         ||||||||||||||    |||||||||||    |||||||||  ||||||  ||||
Db  2129 CTGAAGGTCTTTACGTGGGGTTATAAACATTATCAACACGGTAATATCGAACCTCTCTTCG 2188

Qy  2431 CCTTTGGGCACGGCCTGTCCTCTACACCACCATTTTGCCTTTTCCAAATCTCTCCGTGTCTCA 2488
         |||||||||||||||||||||||||||||||||  |||
Db  2189 CGTTTGGGCCATNGGTTATCCTACAGGTCGTTCAATATATTTCAGAACCTTCGTCTCTCACA 2246
```

```
Matches 1116; Conservative 0; Mismatches 1135; Indels 75; Gaps 9;

Qy  348 ATTGATGTTGAGGCCATCTTGAAGAAGCTACACCCTGGCCGAGAAGGTCGATCTGCTGGCT 407
        |||||| ||||||||||||||||||  |||||||||||||||  |||||| |||||||
Db 81149 ATTGATAAGAAAGCATTGCTCGATAATAATATGACCTTGGCGAGGAGCAGGTCTCGGCTACTTTCC 81090

Qy  408 GGTATCGAACTTCTCTGGCACACAAAAGGCTCTCCCCAAGCATGGAGTCCCCTGCTTT 467
        |||||||||  ||||   |||||||||||||   ||||||||||||||||||||||
Db 81089 GGCGGATACGTTCTGGTCTTTACCGCCCATCGACCGGGCCTAGGGATAGGAGATTGGCCTTG 81030

Qy  468 ACAGATNGGCCCCAACGGCGTAAGAGGGAGGGACCAAGTTCT---TCAATGGCGTCCGGCC 524
        |||||| |||||||||||||||||||||||||||||||||||   ||||||||||  ||||
Db 81029 ACCGACGGCGCCCAAACGGCGTCTTTTGTCGGAGGGTGTGGACTGCAGGG 80970

Qy  525 TGCTTCCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCGTCGAAGAGGCA 584
        ||||||       |||||||||||  |||||||||||||   ||||||||||||
Db 80969 GCTTTCCCGGTTGGGGCATCGCCATCGGAGCGGAGTTGGAACCGGGATTCGGGAAGGAAAC 80910

Qy  585 GGTAAGATGATNGGGAAAAGAGGGCCATCGGTTAAGACTGGTTAAGACTGTCGGCCGACT 644
        ||||||||||   ||||||||  |||||||||||||||||||||||||||||||||
Db 80909 GGCAGTGCCCTAGGTGACGAGGTTCTTTCAAAAGGTGCCCACGTCTTGTTAGCCCCTACC 80850

Qy  645 ATCAACATGCAAACGCTCCCCTCTCGGTGGACGTGGCTTGGCTTGGATTGGATCCG 704
        ||||||       |||||||  |||||||||||||  |||||||||||||||||||
Db 80849 GTTAACATCCAGCGTAGCGTCACAAAACGGCGCAACTTTGAGTGCTTCTCTGAGGATCCG 80790

Qy  705 TTCCTGGGCGGGGCTTGGGACGTGCGGTTCATGACGGGGTACAACTCCATTG 764
        |||||||||||||  ||||||||  ||||||  |||  ||||||| ||||||
Db 80789 ATCCTGACGGCGGGAACTTGCGCCTATATCGAAAGGTGCTCACGAAAJSTCGGT 80730

Qy  765 GCTACGATCAAGCACTTTTGTGCAAGTGATCAGGAGGACAGGGCCCATGATNGGTGCAGAGC 824
        |||||       |||||||||||||||||||||||  |||||||||||||||||||
Db 80729 GCCACGGATAAAAACATTTCGTCGGGAANCAGTGAGATTGAAGCGTACACACCATCTCTTCA 80670

Qy  825 ATCGTGCACGGAGGCGGCTCTCCGGTGAAATCTACGGCCACTCCGGTTCCAGATTGCCA 884
        ||||||||||||||  |||||||||||  ||||||||||||||||||||||  |||||||||
Db 80669 GATATCGGATCGAAACGGACCCACGCGTCGATGGTCTTACCCTTTGAGACGGCGGGGGGTGAAG 80610

Qy  885 GACTCCCACGCCGGTGGCGTTCATGACGGGGTACAACGGCAACGGCATCAAATGGGGTCCGTGCAGC 944
        |||||||||||||  |||||||||||||||||||||||||||||||||||||  |||||
Db 80609 CGGGCAAAGGTCTGGGCGCCGTGAAGTGTCTTGGTATAAACAGGTAACGGTGACTTACACGGCG 80550

Qy  945 GAGAACCCTAAAATATCTTGAGTGGGATGCTTCGAAAGGAATGGGGGTTCTAATC 1004
        ||||||||||||||||||  |||  |||||||  |||||||||||||||||||||||
Db 80549 GAAAGCCCATTGGCTGTTGAACGGAGGTTCTGCGCGGTGACTGGGGTTTTTAACGGCGTGGTG 80490

Qy  1005 ATGAGCGACTGGTACGGCACATACAGTACCACACAGAAGCGCGTTGGCAGGCGTGACCTC 1064.
         |||       |||||||||||||||||||||||||||||||||||  ||||||||||||
Db 80489 ATGTCGGACTGGTTCGGGCTTCGACGGCGCACCACCGGTGAATGCCGTGGGGATCTG 80430

Qy  1065 GAGATGCCCGGACCTCCACGCCTTCCGGAGGAGAAAACACTCAAATTCAACGGA 1124
         ||||||||||||||||  ||||||||||||||||||||||||||||||  ||||
Db 80429 GAGATGCCGGGCCCCGACCTCGGGATCGGCCCGACTCGGAAGGTACTGGCGTCGAAGGCGGC 80370

Qy  1125 AAGCCCCTTTTATCACACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCGTTCAAGAAG 1184
         |||||||||||||||||||  |||||||||||  |||||||||||||||||||||||||
Db 80369 GAGGTTAGTGTCGAGACAGATNCCCGCGCGCTGTGTGCGAATATCCTGACTTTGATTGGAACGC 80310

Qy  1185 TGTGCTGCCTCCGGAGTGACGGAGGAAGCGCCCCGGAGTTTAAGGAGTACGCCATTGATCAACAACACCCCGAA 1244
         ||||||||||||||||||||||||||||||||||||||||||  |||  |||  |||||
Db 80309 ACCGGCGCGTTATCAACGATCATCCGGAGTTAAGGAGTACGCCATTGATCAACAACACCCCGAA 80250

Qy  1245 ACGGCAGTCTCCTCCTCCGGAAGGTTGGCAACGAGGGCATCGTGCTGTCTGAAGAACGAGAAC 1304
         ||||||| |||||||  |||||||||||||||||||  |||  ||| |||  |||||
Db 80249 AGGGCATTGATCAGGCGCGGCGGGGCCGGAAAGCGCAGTTCTGCTACAGAATGACGGGATC 80190

Qy  1305 AACGTTCTGCCCCTTGAGCAAGAAGAAGAGACGCTGATTGTCGGGCCCCAAGCCCAAGCGGGA 1364
         ||||||||||  |||||||||||||||||||||||||||  |||||||||  |||||||
Db 80189 ------CTGCCCATTGGCTCAGCGAAGGTNATGGTCTGCCAATCATCATTGGAACCAAATGGAGCCAAGGTC 80135

Qy  1365 GCCACATACCACGGCGGAGGCTCTGCCCACTCAGGGGCTACTACGGCAGTCACTCCCTTT 1424
         |||      ||||||||||||||||||  |||||||  ||||||||||||||||||||||
Db 80135 GCCCAGGTGATGGGAGGGCGGCTCGGCGCCCAGTTGAACCCTCACTATGTCATCAGTCCTGG 80076
```

```
RESULT 10
REU80928_2/c
WPCOMMENT
Sequence split into 4 fragments   LOCUS REU80928  Accession U80928
       Fragment Name    Begin      End
       REU80928_0           1    110000
       REU80928_1      100001    210000
       REU80928_2      200001    310000
       REU80928_3      300001    371255
Continuation (3 of 4) of REU80928 from base 200001 (U80928 Rhizobium etli strain CFN42 s

Query Match           9.1%;  Score 270; DB 1;
Best Local Similarity  48.0%;  Pred. No. 2.7e-26;  Length 110000;
```

```
Qy    1425 GACGGCCTCAGCAAGCAGCTCGAGACGCCGCCATCGTACACCGTCGGCGCCTACACCACC 1484
            | || ||     |   ||| |    |  |     || |  ||     | || || ||  
Db   80075 CAGGGGCTGGTGGACGCGCTGGGTGAGGAGAATCTGTGCTACGCTCAGGGCTGCAACAAT 80016

Qy    1485 GTTCCTCCCATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGG 1544
             ||         |    |        |  |  ||| ||  |  ||  |   | |   ||
Db   80015 TATC-------GGTTTCAGCCGCTAATCGAAAACCCCACAACCTTCGAGTTTTTCCAAGG 79963

Qy    1545 AGGGTCTTCAACGAGCCCCCTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTC 1604
           | ||     | |  ||   | |||             ||  | | || || |  |    |
Db   79962 AAGGGAGCTAGCTGGCGAACCGGT-----------GAAGGTCGTCGAGGAACCGTCGAGC 79914

Qy    1605 ACCAAGACGGACATGCACCTGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTAC 1664
                        ||| ||  |||| | |   |     | |  |      |   |  |  
Db   79913 CTTGGTGTATGGTTGCCCCCCGTGGCCGAGGGCTTTGTCGATCCCCTTCGC---TTTTCG 79857

Qy    1665 GCCGACATGGAGGGCACGTACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTC 1724
           ||    |||     ||  | ||| |||   || | | ||  |||     |||| ||   |
Db   79856 GCTAGAATGCGCACCATCTTCACAGCCTCGGAAGCCGGCGTCTATCGCGTCGGACTAACC 79797

Qy    1725 GTCTGCGGCACGGCAAAGGCGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAG 1784
                |||     |    ||  || || ||||| |   || || || |||       |    
Db   79796 TCGGCCGGGCTTGGCCGGGTCTATGTGGACGGCAGACTTGTGGTGGACGCTTGGGCTTCC 79737

Qy    1785 CAGGTCCCCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAAT 1844
             | | |  ||     | ||||| |     | |  |   ||||    |||    ||||  
Db   79736 TGGATACGTGGTACCACATTCTTTGAAGAAGGCTGCGAAGAGGTCGTGGGTGAAATCACG 79677

Qy    1845 CTCGTCAAGGGCAACACGTACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACC 1904
           ||||     |||   || ||| || ||     ||||| || |        |  |  || |
Db   79676 CTCGAAGCCGGCCGTACATACGAGGTCGTCGCCGAGTACGCCCGGCACGACCACGTCAAC 79617

Qy    1905 CTCAAGGGCGACACCATCGTCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTC 1964
           ||| |   || | | ||       ||    |||                            
Db   79616 CTCTATATCGCCGCAATTCGGGTAGGAATAGGC------------------------AGA 79581

Qy    1965 ATTGACGACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTC 2024
            ||   | | | || || || |   || | |  ||| ||||     ||   || || |  
Db   79580 TTTTCGGCCGAAGCGGAGATTGCCGAGGCAGCAGCCGTCGCGGCAAAGGCTGATCATGCG 79521

Qy    2025 ATCATCTGCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATG 2084
            |  ||| || ||||   |  | ||||||||||| ||||||||| |||| |      | || 
Db   79520 GTAGTCTTCGTGGGCAGAACGGGCGACTGGGATACCGAGGGCTCCGATCTTCGCGGTATT 79461

Qy    2085 AAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACC 2144
              || |||||  ||| | |||||||  |||  |  ||    || || || || ||||| 
Db   79460 GCACTTCCCGGTCTGCAGAACCAGCTTGTTGAGGCGGTAATTGCGGCCAATCCGAACACG 79401

Qy    2145 GTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCC 2204
            |||||||  | || || ||    || | ||| |||||||||||   ||   | |  ||  
Db   79400 ATCGTCGTACTACAAACCGGTGGACCGGTGGAAATGCCCTGGCTTTCCGGTGCTCGTGCA 79341

Qy    2205 GTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTT 2264
           ||  | ||   ||||||   |||  | ||| | |||||| | || ||||| ||  || | 
Db   79340 GTACTGCAATGCTGGTATCCCGGACAGGAGGCTGGCAACGCGATCGCCGATGTGCTCCTC 79281

Qy    2265 GGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAAC 2324
           ||| |   | | |||||||||   ||| | |  | |||||||  ||||  |   ||||||
Db   79280 GGCAAGGCCGAGCCCTCGGGCCGACTGGCACAGACCTTCCCCGTGCGCTGGGCTGACAAC 79221

Qy    2325 CCCGCGTTTCTCAACTTCCGCACCG------------AGGCCGGGCGCACGCTGTACGGC 2372
           ||| |   |    |      | |||            ||| ||| |    ||  || | |
Db   79220 CCCACACATACTGAGGATGACGCCGTCTATCCAGGTAAGGACGGCCATGTGCGATATGAC 79161

Qy    2373 GAGGACGTCTACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCC 2432
           || |  || | |||||| ||  || |||||||      ||  ||   |       ||||| 
Db   79160 GAAGGTGTGTTCGTCGGTTATCGGCACTACGACCGCCACGGCATTAAGCCGCTGTTCCCG 79101

Qy    2433 TTTGGCCACGGCCTGTCCTACACCACTTTTGCCTTTTCCAATCTCTCCG------TGTCT 2486
           || || || || ||    ||  | |  |||||  | ||  | ||  | |      |  | 
Db   79100 TTCGGTCATGGTCTCGGTTATTCAAGCTTTGCGATGTCGGACCTGACGGTAGGGCTACCA 79041
```

```
Qy    2487 CACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGC 2546
            ||   | ||||  | |||  ||| | ||     |||  |||| | ||     ||| || 
Db   79040 GACGCTGCCGGCGCGGTGACAGTGACACTGGAATTGACCAACATCAGCGAGCGGCCAGGT 78981

Qy    2547 GCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTC 2606
            |    ||||  ||  |||| |||       |  ||||  |    ||   | | || || 
Db   78980 TCGGCAGTGGTGCAAATCTATGTCGGCGATGTTGAAGCATC---TATACCCAGACCGGTT 78924

Qy    2607 AAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCAGCCCGGCGAGA 2652
           || ||||| || | |||| | ||  | |  ||| | ||||| ||||
Db   78923 AAAGAGCTGAAAGCCTTCTCGAAAATTGCCCTGGAACCCGGTGAGA 78878

RESULT 11
AX489445
LOCUS       . AX489445                  2601 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6745 from Patent WO02053728.
ACCESSION   AX489445
VERSION     AX489445.1  GI:22323457
KEYWORDS    .
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6745 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2601
                     /organism="Candida albicans"
                     /db_xref="taxon:5476"
BASE COUNT      917 a    334 c    484 g    866 t
ORIGIN

  Query Match             6.9%;  Score 205.2;  DB 6;  Length 2601;
  Best Local Similarity   46.3%;  Pred. No. 1.1e-17;
  Matches 1178;  Conservative    0;  Mismatches 1253;  Indels  111;  Gaps   11;

Qy     336 ATCATGGCTGATATTGATGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTC 395
           || |  | |    ||||| ||||    || ||     |  | ||  | | |  ||  ||  | 
Db       1 ATGACTGATTCATTTGATATTGATAATATATTATCTCAATTGACATTAGAAGAAAAAATT 60

Qy     396 GATCTGCTGGCTGGTATCGACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCC 455
           |   |  | | ||| || || || ||||| ||| |  |  |  |  |           | |||
Db      61 GGATTAGTAGGTGGAATTGATTTTTGGCATACATATCCAATTTCTCGATTAAACATTCCC 120

Qy     456 TCTCTCCGCTTTACAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCTTCAATGGCGTC 515
               |  | || || ||||| ||||| ||  | ||||| || | ||| ||||||| || 
Db     121 AAAGTGAGATTCACTGATGGTCCCAATGGTATTAGAGGCACAAGGTTTTTCAATGGTGTT 180

Qy     516 CCTGCGGCCTGCTTCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTC 575
           |||  ||| || ||||| || || ||     | | | | | | | || ||  | ||     |  | 
Db     181 CCTCTGGCTTGTTTCCCCTGTGGAACTGGATTAGCTGCTACTTTTGATCATGAATTATTA 240

Qy     576 GAAGAGGCAGGTAAGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTC 635
                  | || ||  | ||| |  ||  || ||| |    || | | |||||| |||||| | 
Db     241 TTGACAACCGGGAAATTAATGAACATTGAAGCTAAATTTAAAAATGCGCACGTGATCTTG 300

Qy     636 GGCCCGACTATCAACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGT 695
           || || || || || || |||   |||  | ||||| || || || || ||  ||  |
Db     301 GGTCCTACGATGAATATTCAACGAGGCCCATTAGGTGGCCGAGGATTTGAATCATTTTCT 360

Qy     696 GAGGATCCGTTCCTGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTC---AGAGC 752
           || ||||| |   |  | ||       ||    ||| | ||    || ||||   |    
Db     361 GAAGATCCTTATTTAACTGGACAAATTGCCAGTGCTATAATTAAGGGTATTCAATATGAT 420

Qy     753 ACTGGAGTGCAGGCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATG 812
           | || | |    || ||| | || || | | | || ||||||   || ||     | |  
Db     421 AATGAAATTGGTGCCACGGTTAAACATTATGTTTGTAATGATTTAGAAGATGAAAGGAGT 480

Qy     813 ATGGTGCAGAGCATCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAG 872
```

```
Db    481  GCTAGTGATTCTTTGGTTACCAAGAGCATTGAGAGAAATTTATTAGAACCATTAGA                      540
           ||  || |||                                              ||  ||

Qy    873  ATTGCTGTGCGAGACTTCCCAGCCGGGTGCGTTCATGACGGGATCAATGGC                                932
           || |||| |                                         ||||||

Db    541  ATAGCTATTAAAAGAAAAGTAATCCAATTTGTTTGATGACAGGGTTAATAAAAGTTAATGGT                     600
           ||||| |                                               || |

Qy    933  GTGTCGTGCAGGGAGAAGACCCTAAATATCTTGAGGGAGTGCTTCGAAGGGAATGGGGTTGG                     992
           ||  |||                                                ||||

Db    601  GAACATGTTTCTCAAAGTAAATTTTTTGCAAAATAATTATTACGTGATGATGGAATTGG                        660
           ||                                                     |||

Qy    993  GATGGCCTAATCATGAGCGACTGGGTACGGCACATACAGTACCACAGAGCCGTTGTTGGCA                       1052
           ||                                                     ||

Db    661  CAAGGGACAAATTATATCTTGATTGGGGACAATATACAGTACCATTGAAAT                                720

Qy   1053  GGCCTCGACCTCGAGATGCCCGGACCTCCACGCCGCTTCCGGAGGGAGAAACAACTCAAGTTCAAC                  1112
           |||                                                    ||||

Db    721  GGTTTGGATTTAGAAATGCCAGGGATCTCCAATTTCCGTAATAACAATTGTTGACTTAGT                       780

Qy   1113  GTCTCCAACGGGAAAGCCCTT---                                                           1169
           |||||

Db    781  ATGATTAAATCAAAAGAAATTGCATATATCAAACATTGGATGATCGTGTACACGTTGTTTTA                     840

Qy   1170  CAGTTGCTCAAG------                                                                 1223
           |||||

Db    841  AAATTAATTAAATTTGCCAAAACAGTCAAGTGTTGTGTCGGTCACTGAAGATTTGACTTAGT                     900

Qy   1224  ACTGTCAACAAACCCCGAAAACGGGCAGCTTCCTCCTCGGGAAGGTTGGCAAGGGGCATC                       1283

Db    901  TCCAAGTCTATTGCAATAATTGGTTCCTAATGCAAAAATTGTTGCCGGCTTATTCAGGGAGGAGGG                  960

Qy   1284  GTGCTGCTGAAGGAACGAGAAACAACGTTCTGCCCTTGAGCAAGAAG-------                             1328

Db    961  GTTTTATTGAAAAATGATAAATAAATATATTTATTACCATTGAAAAGGGATGACGTCGAACTTCAGC                1020

Qy   1329  ------AAGAAGACCTGATTGGTCGGCCCCAAGCAGCGCCACATAGGATGACGCTCGACTTCAGC                  1385

Db   1021  TCCAAGTCTATTCCAATAATTGGTCTTCCTAATGCAAAAAATTGCCGCTTATTCAGTGGTTGGGG                  1080

Qy   1386  TCTGCCCGACTCAGGGGCCTACTACGGTCACTCACTCCCTTGACCGGCCTCAGCAAGCAGCTC                    1445

Db   1081  TCAGCCACTTTACCAGCTTATTACACCACTACTCTCCTTATAATGCCATTGTTGAGAAATTA                     1140

Qy   1446  GAGACGCGCCCATCGTACACGTCGGCGCCCTACACCACACGGTTCCT------                              1490

Db   1141  TCTTCTATTTCCAAATTTGATATTACATCAAATATACTATTGGTGCAAAAGCA                              1200

Qy   1491  --------CCCATTCTAGGCGACGACAGTGCCTCACGCGCGCTCCGGCATG                                1538

Db   1201  TATAAAAATTTTACCAGAATTAGTCCACAAGTGATCAATCCCAAAACTGGTAAAACCCGGC                      1260

Qy   1539  CGCTGGGAGGGTCTTCAAGGAGCCCCTGGTACCCCT----AACCGGCCAGCACATT                          1589

Db   1261  TTTTCCATGAAAATTTTATAAAAGCCAAAATCCGTTCCTAATGAAAATGAGAAATTATTT                       1320

Qy   1590  GACGAGCTCTCTTCTTCACCGACGACAGCACTGCTGGGGACTACTGTAAAATCGAGAATATC                     1649

Db   1321  GATGAAATTAGATACAGAAATATCTGAAGTGATTAATTTAGGTGATTATTATCATAAAGATATC                   1380

Qy   1650  GCAGACACGTGG-----                                                                  1703

Db   1381  CCTTCTAATGGGTTTATAATTGAATTGTCACACCATCAAAAACCAA                                     1440

Qy   1704  ACCTACGAGCTCGGCCTCGTCGTCTGGGGACAGCGCAAAGGGGTACGGTAGAGACCAGCTC                      1763

Db   1441  CATTATGAATTTGGTTTAACTGTTCATTGTTCAGCACAAATATATTTTTGAATTCTGGAACCATT                  1500

Qy   1764  GTCGGTCGACAACGCCCACCCAAGCAGGTCCCGCGGATCGCCTTCTTCGGCTCCGCCACCCGC                    1823

Db   1501  GTTGTTTGATAATAAAAACCAAACAAATTAAACTGTTCATTGGTTAACGTTAGATTAAACTT                     1560

Qy   1824  GAGGAGAGACGGGCCGCCATCAAACTCGTCAAGGGCCAACACGTACAAGTTCAAGGATCGAGTTC                  1883

Db   1501  GTTGTTTGATAATAAAAACCAAACAAATTAAACTGTTCATTGGTTAACGTTAGATTAAACTT                     1560
```

```
Db   1561  GAAGAACGAGGAGATCAATTGAATTACATCAAGGTAAAAACTTATAAAATCATTGTTGAATAT              1620

Qy   1884  GGCTTCCGGCACCACCTTACCACCCTCAAGGGGGACCACCATCGTCCCGGGCCACGGGCTCCCTC           1943

Db   1621  GGTTCGGCACCTACATTTACATTTAAAAAGATCAAGTTGGAGAATATTTTGGTGTGGGATT               1680

Qy   1944  CGCGTCGGGCGGCTGCAAGGTCAT---TGACGACCAGGCGGACGAAATCGGATCATCAGTCAAAT          2000

Db   1681  CGTCTTGGTATGAGAAGTAAATTATTAATGATGAACAAGAAAATCATCAATGCAGTAAAT               1740

Qy   2001  CTCGCCAAGGAGGACCACGGACCAGGTCATCATCACATCTGCGCGGGCCTTACGCCGACTGGGAGACC        2060

Db   1741  TTGGCTAAAATCAGTAGATTACAGTTAGTGATTTTAATTATTGGATTAAATAAATAAAGATTGGGAAAGT      1800

Qy   2061  GAGGGCGGGCCGAGGCAGCATGAAGAGCTCCCCGGCCGTGCGTGGGACTCGATCAGTCAAGTTCAAC        2120

Db   1801  GAATCTTATGATCCGTTCCTGAATTTAGTGGAATTACCTGGGATTAAATTGATTGAATCG               1860

Qy   2121  GTGGGGCCCGGGCGGCGGAAACCCAAAACCGTCGTCGGTCCGTCATCGCCGAGGACGGGGCACCCCCGAGGGAGATG  2180

Db   1861  GTAATTGGACGTTAAATCCAAAACAACCATTATTGTCAAATCCGGTACACGTTGAATTT               1920

Qy   2181  CC----CTGGCTCGACGGCACGCCCGTTCGCCAGGGACAACCCGCCGTTCTCAACTTCCGACCACCGGGCGGGG   2237

Db   1921  TCTGAATTGGTTGGATAAAACTGAAAGCTTTGTTGTTGCATTGTTGGTATGGTGGGAAATGAAAGT         1980

Qy   2238  CGCAACTTCCATTGCCGACGTCGTCTTTTGGGGCACTACAACCCCTCGGGGCAAGCTCGTCCCTC         2297

Db   1981  GGGAATGGCCATTTGCCTGCGATATTTTATTTTGGTAATGTCAAATCCTTAATGGGAAAATATCATTG      2040

Qy   2298  AGCTTCCCCAAGGCGCCCTTGCAGGACAACCCCGGCGTTTCTCAACTTCCGACCACCGAGGGCCGGG        2357

Db   2041  ACTTTCCCATTGAGAAAAAATAATGATAATCCAACACTTTTTTTGAATTTTTAAAACTGGAAAGGT        2100

Qy   2358  CGCACGGCTGTACGGCGAGGACGTCTACAGTCGGGGTACAGGTACTACGGAGTTTGGCCGACAAG        2417

Db   2101  CGAGTATTATAATGGGAGAAGAGATATTTTTGTTGGTACAACGGGTACGGTCCGGGGCATG            2160

Qy   2418  GACGTCAAATTTCCCCTTTGGGCCCACGGCCGCTGCCTCACCACCACTTTTCCCTTTTCCAATCTC        2477

Db   2161  CAAGTTGCTTTCCCATTTGGATTTGGATTATCATAATACTAAATTTGGAATTTTCAGATTTG           2220

Qy   2478  TCCGTGTCTCACAAG----                                                        2507

Db   2221  AAAATTGGAGAGAGAACAAAAGATAAAAGAAGAAGAGAAGAGAAGAAGAAAGTCTTGTT              2280

Qy   2508  GTGTCGTCCTCTCCGTGAGAAAGAAACACCGGCGTTCCGTGTCGGCGGCCCAGCTCTAC            2567

Db   2281  GTTTCATTGTCGGGTGAAAAATTATGGTAAGATCTTGAAGTGATTCAAGTTTAT               2340

Qy   2568  GTCAAGCCCCTCCAAACGGGCCAAGGAATTAACCGCCCCGTCGAAGGAGTCAAGGCTTCGCA            2627

Db   2341  GTATCTAAAGTCGGACATCTGATCTGATTTA---GGCCAATTAAAGAAATTGAAAGGTTTTTTACT      2397

Qy   2628  AAGGTCGAACTGCAGCCCGGGGAGGACGAAAGGGGGGTGACAATCGACGAGAAGTAC             2687

Db   2398  AAAGTTAATTACACAACAAAATGTAAATAGGTCAGTGGTGTCGAAAAGGTTGACTATGAG          2457

Qy   2688  GTCGGCTGCGTGTATTTTGATGAGGAGCCGGGAACGGGAGCGGCTGTCGAAAGATTGAG           2747

Db   2458  TCAGTTTCGTTTTTCGAATGAATATGCCAAATCAAATGGTTCAACTGGCCCAATACAAA          2517

Qy   2748  GTTATCGTTGTAGCGGACGACAGCAGCGG                                             2769

Db   2518  GTTCATGTTGGTAAATAGTAGTG                                                   2539
```

KEYWORDS   beta-glucosidase; elongation factor G; fusA; succ; succinyl-coa
              synthetase alpha chain; succinyl-coa synthetase beta chain; sucD;
              TPP-requiring enzyme; transfer-RNA-Leu; TTA Leu codon.
SOURCE   Streptomyces coelicolor A3(2).
  ORGANISM   Streptomyces coelicolor A3(2)
              Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
              Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1 (bases 1 to 37445)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
              Kinashi,H. and Hopwood,D.A.
  TITLE   A set of ordered cosmids and a detailed genetic and physical map
              for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
   MEDLINE   97000351
   PUBMED   8843436
REFERENCE   2 (bases 1 to 37445)
  AUTHORS   Seeger,K.J. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 37445)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-JUL-1998) Streptomyces coelicolor sequencing project,
              Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
              CB10 1SA, E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
              David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
              Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT   On May 9, 2002 this sequence version replaced gi:3288600.
              Notes:
              Streptomyces coelicolor sequencing at The Sanger Centre is funded
              by the BBSRC.
              Details of S. coelicolor sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
              numbered using the following system eg SC7B7.01c. SC (S.
              coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
              strand).
              The more significant matches with motifs in the PROSITE database
              are also included but some of these may be fortuitous. The length
              in codons is given for each CDS.
              Usually the highest scoring match found by fasta -o is given for
              CDS which show significant similarity to other CDS in the database.
              The position of possible ribosome binding site sequences are given
              where these have been used to deduce the initiation codon. Gene
              prediction is based on positional base preference in codons using a
              specially developed Hidden Markov Model (Krogh et al., Nucleic
              Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
              of Bibb et al., Gene 30:157-66(1984) as implemented at.
              http://www.nih.go.jp/
              jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
              correct initiation codon. Where possible we choose an initiation
              codon (atg, gtg, ttg or (att)) which is preceded by an upstream
              ribosome binding site sequence (optimally 5-13bp before the
              initiation codon). If this cannot be identified we choose the most
              upstream initiation codon.
              IMPORTANT: This sequence MAY NOT be the entire insert of the
              sequenced clone.   It may be shorter because we only sequence
              overlapping sections once, or longer, because we arrange for a
              small overlap between neighbouring submissions. Cosmid 8A6 lies
              between 3F9 and 1F2 on the AseI-A genomic restriction fragment.
FEATURES   Location/Qualifiers
     source   1..37445
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 8A6"
     misc_feature   2..107
                     /note="nominal overlap with St8A6 from 1 to 106
                     nominal overlap with cosmid 3F9 from 19725 to 19830"
     stem_loop   657..711
                     /note="hairpin loop with 23 bp stem"
     gene   complement(730..2160)
                     /gene="SCO6581"
     CDS   complement(730..2160)

     gene   complement(730..2160)
                     /gene="SCO6581"
                     /note="SC8A6.02c, probable transmembrane transport
                     protein, len: 476 aa; some similarity to e.g. TR:Q51330
                     (EMBL:U04007 5) oxalate:formate antiport protein from
                     Oxalobacter formig enes (418 aa), Opt: 226
                     z-score: 230.1 E(): 1 .4e-05, 24.1% identity in 436 aa
                     overlap"
                     /db_xref="SPTREMBL:O87836"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transmembrane transport protein"
                     /protein_id="CAA19774.1"
                     /db_xref="GI:3288602"
                     /translation="MQQCALARSFDPARGAIDMTTTDVTRVAAYREVTDRNGRVYRGE
                     SDIDLMGRKRKWMVILPWIGNMGISSAEYAFAASAEDTLHEAQHWSSGSIYKWMTAWVF
                     CQAAVAFPAGRLRESGKLPARWAMMLGSAGTLLGYILSLAFAPHVSLAFIGFGVFSGMG
                     AGMVTATCVNNWGKWPERRGGKTGFVNGGPAYGSVPFVFIFHGTMDTSNFRWLVSA
                     GVFLAAMVAIAGFYFRDPPKNWWPAAVDPLNPPADPRARRSLEKNPPAVKQYSPMEAW
                     KTGRVALMWFCLLACTSGVNIFGIAFQVDIGEEAGFAAGVYAAAMSLKAIVNGTRGVI
                     GWLSDIYGRKQCLLTVCAILGLAQPGIIWAAEIKNLPLFIVFSAISGFGGGAIFPMFA
                     ALTADYFGENNAINTNYGMVYSSKLVSGLGAGMGSVVVGVWGVWGYNGAFSLAGSISIFAGF
                     VALSLRPPGRPEPGKRVTPNPQPLGEG"
     gene   complement(2233..4377)
                     /gene="SCO6582"
     CDS   complement(2233..4377)
                     /gene="SCO6582"
                     /note="SC8A6.03c, unknown, len: 714 aa; similar to many
                     hypothetical proteins including several from each of the
                     arche ons Archaeoglobus fulgidus TR:O28341 (EMBL:AE000970)
                     conser ved hypothetical protein AF1938(673 aa), fasta
                     scores; Opt: 960 z-score: 1289.8 E(): 0, 35.6% identity in
                     693 aa overl ap and Pyrococcus horikoshii TR:028624
                     (EMBL:AB009518) hypot hetical protein PH0IV016 (460 aa),
                     fasta scores; Opt: 1280 z -score: 1246.9 E(): 0, 44.3%
                     identity in 461 aa overlap. Co ntains PS00216 Sugar
                     transport proteins signature 1. Also s imilar to S.
                     coelicolor hypothetical protein SC9B10.09 (E() : 6.1e-15,
                     33.9% identity in 375 aa overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAA19775.1"
                     /db_xref="GI:3288603"
                     /translation="MAEDRIARVRGLLDAVRGEGRTALTAPEGKVLADAYGIAVPGEE
                     LARDVDEAVACAARFGGPVVMKIVSPDILHKTDAGGVVGVGAEEAVAAAFCRIVDNA
                     RAYDASARIEGVQIOELLEPQGQEVIVGAVTDPFEGKVVAFGLGGLVEVLKDVTFRLA
                     PVDADEALSMLDSIRAAEVLRGVRGQAGVDRWAVAEQIRRVSELVADPEIAEVDLNP
                     GDDAVYDDILRQAGVIRAPGLNEMLEYARALPVLPAPEGDNVVITGAGGSGVLLSDA
                     VTDNGLSLMEIPPDLDRAFRAFIPFGAAGNPVDITGGEPPSTYEAFIRLGEDPRIH
                     ALVLGYWHHIVTPPMVFAELTARVVAEERGRGVEKPVVASLAGDVEVEEACQYLFERG
                     GNSYWRNLVDGFAGEIHPVNPKADDILGRKAYKSVTDPGEVDVAVFAIPARFVAAA
                     LEEVGRKIRFPNAVLIPSGFAETGEQALQDEIVAVAERHGVRLLGPNIIGIYYSTWQDLIC
                     ATFCTPDVKGGVALTSQSGGIGMAILGFARTYKTGVSAIVGLGLLVELKDVTFRLA
                     GEDPHTECTAMHLEDLKDGRAFVEAARATVPRKPVVLKAGRTAAGKAAGSHTGALA
                     WAYPYTTEKPVAVLGAKYRWARAAGLLGGGS"
     gene   complement(4382..5614)
                     /gene="SCO6583"
     CDS   complement(4382..5614)
                     /gene="SCO6583"
                     /note="SC8A6.04c, probable transferase, len: 410 aa;
                     highly similar to TR:O06644 (EMBL:U82167) formyl-CoA
                     transferase from Oxalobacter formigenes (428 aa), fasta
                     scores; opt: 95 4 z-score: 1732.0 E(): 0, 50.7% identity
                     in 428 aa overlap. Also similar to many dehydratases e.g.
                     CAIB_ECOLI L-carnit ine dehydratase (EC 4.2.1.89) (405
                     aa), fasta scores; opt: 285 z-score: 317.7 E(): 1.9e-10,
                     26.2% identity in 427 aa o verlap"
                     /codon_start=1
     misc_feature
                     /note="Ps00216 Sugar transport proteins signature 1"

```
                    /transl_table=11
                    /product="putative transferase"
                    /protein_id="CAA19776.1"
                    /db_xref="GI:3288604"
                    /translation="MTPAKALEGIRVLDMTHVQSGPSATQLLAWLGADVVKLEAPHGDI
                    TRGQLRDLPDVDSLYIFMLNCNKRSITPLNTKSERGKEILTELIRRSDVWENFNGPGAV
                    DRWGFTWDRVKEINPRIVYAISIKGFGEGPYYTAFKAYEVAAQAMGSMSTGFEDGPPL
                    ATGAQIGDSGTGVHVHVAGILAALYQREHTGRGQRVNVAMQHAVLNLCRVKLRDQORLS
                    HGPLAEYPNEDGEVPRSGNASGGGPGWAVKCAPGGPNDYVYIVIVQPVGWQPLSEL
                    IGRPELAEDPEWATPRARLPLKLINKMFQLIEEWSSTLPKWEVERLNAHNIPFCGPILST
                    KEIIEDDSLVANEMVVPHPERGEFVTVGSPLKLSDSPVEVTSSPLLGEHNEEVVVG
                    ELGLGDEELRLLKSSGVI"

RBS          complement(5623..5627)
             /note="possible RBS upstream of SC8A6.04c"
gene         complement(5626..7308)
             /gene="SC06584"
CDS          complement(5626..7308)
             /gene="SC06584"
             /note="SC8A6.05c, probable TPP-requiring enzyme, len: 560
             a.a; similar to many e.g. ILVB_MYCTU probable acetolactate
             synthase (EC 4.1.3.18) (547 aa), fasta scores; opt: 733
             z-sco re: 1305.6 E(): 0, 36.4% identity in 557 aa overlap.
             Contai ns PS00187 Thiamine pyrophosphate enzymes signature
             and Pfam match to entry PF00205 TPP_enzymes, Thiamine
             pyrophosphate enzymes, score 343.90, E-value 5e-106"
             /codon_start=1
             /transl_table=11
             /product="putative TPP-requiring enzyme"
             /protein_id="CAA19777.1"
             /db_xref="GI:3288605"
             /db_xref="SPTREMBL:O87839"
             /translation="MPDDTQDVLSGGHLVAKALKAEGVDRIYTLCGGHIIDIYDGCVD
             EGIEVDVRHEQVAAHAHADGYARITGKPEGCAVVTAGPGTTDAVTGVANAFRAESPMLL
             IGGQGAHTDQHKMGSLQDLPHVDMMTPITKFAATVPDTRARADNWSMAFRECYHGAPGP
             SFLEIPRDVLDAKVPAEKARVPRRAGGYRASTRSAGDPEAVERLADLVHAEKPAVLLG
             SQVWTTRATEAAVELVRTLNVPAYMNGAGRGTLPPGDPHHFQLSRYAFSNADIVIVIV
             GTPTPDFRMGYKRLSPDATVVQIDLDYRVGKNRDIDLGIVGDBAGILVLKSVTEAASGR
             INGAAKRKEWLDELRAAERTALDKRLPQLRSDASPIHPYRLVSEINDFLTEDSIYIG
             DGGDIVTESGQVVOPKSPGHWMMDPGFLGTLLGVGVPFVLAAKQARPDKEVVALFGDGAF
             SLTGWDFEETLVRDLPFVGIVGNNSSMNQIRVGQAVKYGKERERVGNTLGDVPYDEFA
             RMLGGHGEEVRDPADIGPALRRARESGKPSLINVWVDPDAYAPGTMNQTMYK"
misc_feature complement(5683..7254)
             /gene="SC06584"
             /note="Pfam match to entry PF00205 TPP_enzymes, Thiamine
             py rophosphate enzymes, score 343.90, E-value 5e-106"
misc_feature complement(5962..6021)
             /gene="SC06584"
             /note="PS00187 Thiamine pyrophosphate enzymes signature"
             complement(7316..7319)
```

Query Match   6.6%; Score 197.4; DB 1; Length 37445;
Best Local Similarity 47.1%;
Matches 1152; Conservative   0; Pred. No. 9.1e-17;
Matches 1152; Conservative   0; Mismatches 1176; Indels 119; Gaps 13;

```
Qy    372 AAGCTCACCCTGGCCCGAGAGGTCGATCTCGTGGCTGGTATCGACTTCTGGCACACAAAG 431
          |||||| ||||| |   ||| ||| ||||||| || || |||||||| ||| || |||||
Db  27611 AAGCTGACACGCGCCCCGGCACTGCGTGGACGGGCCACCGACCTGGCCACGAGA 27670

Qy    432 GCTCTCCCAAGCATGGAGAGTCCCCCTCCTCCGGCGTTTACAGATGGCCCCAACGGGTAAGA 491
          ||||| |  | | |||| || ||  |  || || ||| |||| | ||  | |||||||||
Db  27671 GCGGAACCACGCGGTGGAGCTGAGGGAGTTGGTGATGTCGGACGGGTCCGGCGGGCGTACGG 27730

Qy    492 GGGACCAAAGTTCTTCAATGGGGTCCCTGCGGCGCCTGCTTCCCTTGCGGCCACGTCGGTCGGT 551
          ||| | |   || ||  || || |||||| ||  | ||| ||| ||  ||||| |||||
Db  27731 GGCGAGGCCTGGGCGACGAGCACCTCTCTCCTGCCCTGCGTGGGTCGGCGGGCGCATC 27790

Qy    552 TCCACATTCACCAAAACTCTGCTCGAAGAGGCAGGTAAGGATGAGTGGGCAAAGAGGCCATC 611
          || |    |||  |  | |  |  || |  | || ||  |   ||| |||  ||| || |||||
Db  27791 GCCACCTTGGGACGAGAGGCGCTGGTCGGACGAAGAACTCGGTGCGGCCCACTACGTGGCCTA 27850

Qy    612 GCTAAGAGTGCGCCATGTGATCCTCGGCCCCGACTATCAAACATGCAACGCTCCCCTCTCGGT 671
          ||| | | || | || ||  | | |||| || | |||  |  ||| ||  || | |  | ||
Db  27851 CGCAAGGGCGTGGAGCTTGGACGTTCCTCCGCCCCACCCTGCAGCCCGGCTGGGC 27910

Db  27911 GGGCGGCCACTTCGAGTCCCTCTCCGAGGACCCGGCCGCGGATCGGCGGCGGCCG 27970

Qy    672 GGACGTGGCTTCGAGTCGGTGATTGGTGAGGATCCGGTTCCTGGGCGGCGGCTGGGC 731
          ||||| || ||| || | || |  ||| || |  |||  ||  ||| |||| |||
Db  27971 CTGGTCCCGGGGGATTCCAGGGCCAGGGGCGGCCGCAGGCGCACTACGTGGCCAAC 28030

Qy    732 CTCATCCGCGGGCGATTCAGAGACACTGGAGCTGCAGGGCGGATCAAGCGGGGCT 791
          || ||| ||| ||  |||| | |  || |||  || || || ||  | |  ||||||
Db  28031 GACTCCGAGACCGACCGAGCCCCTCACCGTCGACGTGGGGGGAACGGCGGCGGCTCCGTGAA 28090

Qy    792 GATCAGGAGGACAGGCGCCATGAATGGTTCAGAGACCATGCTCACGGAGGACGAGGTGGGC 851
          |   |||| |||| |||   | | || || ||| ||| |  |||  || || ||   ||||
Db  28091 GTCTACCTTCGCCCCCCTTCGGACGGCGGTG---GCCCCGCGGGGTCCGGCTCGTCAATGGCG 28147

Qy    852 ATCTACGGCACTCCCGTTCCACAGATTGCTGTGCGAGACTCCCAGCCGGCGGGTGCGTTCATGACG 911
          |||||      ||||    |||  ||| | |  || |||  | |  | ||| |||  |  |  |
Db  28148 GGGTACAACGCGGGGTCAACGGGGTTGGGATGGCCACCACGATGACGGACGACTGGTACGGCACATACAGT 28204

Qy    912 GCGTACAATGGCATCAATAGGCGGTGTCGTGCAGGGAGAGAACCCTAAATATCTTGATGGGATG 971
          ||| || | ||| |||  |||| | || ||  ||  || ||| |||   ||| ||  |||
Db  28205 CTGAAGAGCGAGTGGGGCGGCTTGCAGGTCCTCAAGTCGGTCACGGAAGCCGCGTCGGGC 28264

Qy    972 CTTCGAAAGGAATGGGGTTGGGATGGCCACCACGATGACGGCGAACGC--CCTCCTCACCGACCCG 1031
          |||  ||||| | |||  |  | || |||||  | || ||| ||||   |  || || ||  |||
Db  28265 ACGACGCCGCACCCGCCGCCGCCTGGGCGTCTTCGACCTGGCCGTGCTGGCGACCTCCAGGGTTCCGA 28324

Qy   1032 ACCACGAGAGACCGTTGTGGCAGGCCCTCGACCTGGCCGTGCTGGCGACCTCCAGGGTTCCGA 1091
          ||||| || ||| ||  |  |||  ||  | |||  || ||| || ||||| ||  || |||
Db  28325 GCGGAGGCGGCCGTTGGCCCGCGGTGGGCCGAGGACCGGCTCCGAGAAGTGTGCTGCCTCCGGAGTGACGGAGGAAC 28384

Qy   1092 GGAGAAAACACTCAAGTTCAAACGTCTCCAACGGGAAAGCCCTTTATCCACGTCATTGACCAG 1151
          ||| || |||| || || ||  ||||  |  | |||  |||  |  ||  ||| || ||||||
Db  28385 AAGCCACGGCGCCGGCCGGACTGCGGGCGGGGCGGGCCGCGGGCCGACGTG 28444

Qy   1152 AGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCTGCCTCCGGAGTGACGGAGGAAC 1211
          || ||  |  || ||||| || || ||| ||  | || || ||  |||| || |||
Db  28445 TCCCCGGTCGCGGTGTCCGGCCAGGCCAGGGCCGAGCTGGGGGTCGGAGGT 28504

Qy   1212 GGCCCCGAGACGACGTGTCAACACACCCCCGAAAACGAGGCAGCTCCTCCTCCGGAAGGTTGGC 1271
          ||   || | |   || | | || ||  ||  ||||| | | ||| ||| || |||| || |
Db  28505 GGGGGGGCGGCTGGACGGGCGGGGCGGGGCGGAAGCCTTGAGCAAGAAGAAG 28564

Qy   1272 AACGAGGGGCATGCGTGCTGCTGAAGAACCAGAACAAACGGTTCTGCCCTTGAGCAAGAAGAAG 1331
          ||  || || ||  || ||  || |   | || || || |  | ||  |||  || ||| |||
Db  28565 GGTGGCGAAGCCGTCCGGGCGGGCCCCCACGCGGCCGGCCGGGGACTGGCCGCT 28624

Qy   1332 AAGACGCTGATTGTGGCGCCCAACGGCCAAGCCAGGCCCACATACCA------------ 1375
          ||  || || |  || | ||  || |   ||||| || |  | |||
Db  28625 CGTGCGGCCGTCGCGGGCCGGGCCCCGTGCTCGGCGGCCGGGGGGCCGGGGGCCGGCGGAC 28684

Qy   1376 -----CGGGCGGAGGGCTCTGCCGCACTCAGGGCCGCCACTAGGCAGTTC------------ 1415
             |||||| ||  || || ||| ||||| ||| || | |||  | ||
Db  28685 GAGCGCACCTCGGGAGCACGGTCGCCGGCCGGGGGAGCAGGTGCCCGAGAGTGACGGGCTCCG 28744

Qy   1416 --ACTCCCTTTGACGGCGCCTGACGAAGCAGCTCG-AGACGCCGGCCATCGTACACGGTCGGC 1472
            || || || |||  || || || || |  |  ||| | |||  || | || ||||
Db  28745 GCGGCGCAGCCGCGGCGCGTCTTCCCGGGGGAGTGGGAGGCCGGCTGGACGAGGGCTACGGGC 28804

Qy   1473 GCCTACACCACCGGTTCCTCCCCATTCTCAGCGGCAGTGCCTCACGCCCGACAGGGCTCG 1532
          || |   | |   | |  |  ||  | | ||||| || || ||  | |   | |||| |
Db  28805 GAACTGCGCGGCGGGCGTCGTGCACGTTCCGGTGACTACTACCACCCACCAAGGCGG 28864

Qy   1533 GGCATGCGCGTGGAGGGTCTTCAACGAGCCC--CCTGGGTACCCGTTAACCGCCTAAACGGCACATTG 1590
          || |  |||||||||| || ||| |||||    ||||| ||| || || ||||| | |   ||||
Db  28865 CCCCCACTGGACCCGGACACATGCACCGACCCGTGCGCCGGCCGGGCCGTGTCGTTCTGCTG 28924

Qy   1591 ACGAGCTCTTCTTCACCAAGACGACGATGACCTGGTGGACTACTACCACCCACCAAGGCGG 1650
          | || |  | ||| || ||  || || || ||  |  ||| | || || ||| ||||| ||
Db  28925 CGGATGCTTCGACGCGGACGGCGACGAGCTGTTACGGCGGAACGGCGACGGCGGCGGCGGCCTG 28984

Qy   1651 CAGACACGTTGGTTACG--CCGACATGGAGGGGCCACGTTACACCGCGGGCGAGAGAGACTACGG 1708
          || ||  | |||  | |   |||| || |  |||| ||| ||| || ||||||||||  | | ||
Db  28925 CGGATGCTTCGACGCGGACGGCGACGAGCTGTTACGGCGGAACGGCGACGGCGGCGGCGGCCTG 28984
```

```
RESULT 13
LOCUS       AR173232                3241 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 11 from patent US 6303767.
ACCESSION   AR173232
VERSION     AR173232.1  GI:17912723
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3241)
  AUTHORS   Betlach,M.C. and McDaniel,R.
  TITLE     Nucleic acids encoding narbonolide polyketide synthase enzymes from
            streptomyces narbonensis
  JOURNAL   Patent: US 6303767-A 11 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1..3241
                     /organism="unknown"
BASE COUNT      520 a   1174 c   1125 g    422 t
ORIGIN
```

```
Query Match     6.2%;  Score 184.4;  DB 6;  Length 3241;
Best Local Similarity   48.2%;  Pred. No. 5.8e-15;
Matches 1028;  Conservative    0;  Mismatches  956;  Indels 150;  Gaps 12;
```

```
Qy    1709 CGAGCTCGGCGGCCTCGTCGTCGTCGCGGCGACGGCGAAAGCGGTAGTAGGACGAGCTCGTCGT 1768
                |  ||| |||||      || | ||  ||| ||||     ||||  |  ||||  |||| ||
Db   28985 CTGGAGC-CCCGCGTTGGTGCGCCGCGCACACCGTCAGATCCCGCGCCGCTGTGTCC 29043

Qy    1769 CGACAACGGCCACCAAGCAGAGTTCCCCGGCGGGGGGATGGCCTTCTTCTGGCGTCCGCCAGGA 1828
                |||| ||  || || ||| ||||||    ||| |||| | ||| | ||||| |||||
Db   29044 CCGCACGCCGGCGGCGCTCCGTGGTCCCTGGGGCGTTGCGGCGGTGAGCCTGAACGAC 29103

Qy    1829 GACGGGGCCGGCCGATCAATCTCGTCAAGGGCAACACG------------------ 1862
               ||||| |  ||| |||      ||| ||| | ||
Db   29104 GGACGGGACGACACCCTGCTGGAGGGGGACTTCCCGCCGTCCACCGAGGATCCGGCCGGTGAT 29163

Qy    1863 ------------------TACAAGTTCAAGATCGAGTTCGGCTCCGGCCGACCCAACCTACACCCT 1906
                              |||  |||| || | ||||| ||| || |||| ||  ||||||
Db   29164 GCACGTCAACCGCCCGCCCAGTACGCCACCGGCCACCCCAGTACGTCCTCGTTCGCGGGC 29223

Qy    1907 CAAGGGCGACA------CCATCGTCCCGCGGCCCACGGCTCCTCTCCGCGTGCGGCGGG 1954
               |||||  ||        | |||||  ||||| |||||  |||||||||||| ||  |
Db   29224 GCTGGGTGCCCGGCGGCGAGCTGGCACCCGGCCACCCCGACTCCTCGTTCGCGGGC 29283

Qy    1955 CTGCAAGGTCATTGACGACCAGGCCGGAAAATCGAAAAGTCCGTCGCCCTCGCCCAAGGAGCA 2014
                |||   ||     |||   ||| || |  ||||||  | ||| || |  |||  |
Db   29284 CCCGCCGCCCCGACGTGACCGGTGCTGCGCCGGCTGTCCCCGGGCGTTGCCGGC 29343

Qy    2015 CGACCAGGTCATCATCTGCGCGGGCCCTTAACGCGCACGTGGGAGGCGGCGCGCCGACCG 2074
               ||| || | |   | |||| |||||||   || | ||  ||| |||||  || |||
Db   29344 GGACGGCGCGTCGTCGCGGCACCACCGGGGAGTGCGGGGCTACGACCG 29403

Qy    2075 CGCGAGGCATGAAGCTCCCCGGCCGGCGTGGACCAGCTCATTGGCCGGCGCGCCCGAA 2134
              |||| |      |  | ||||||| ||| || | |||      ||| ||| |||||
Db   29404 TACGGACTGGCGCTCTCGGGCGCCACCACCAGGAGTCCGGCCGGTCCGGCGCCCAA 29463

Qy    2135 CCCAAACACCTCGTCGTTGTGCTCATGCAGGAGGATGCCCGCGGCGGCGTCTGACGC 2194
              ||||| |  ||||| || |||||||  |||| || ||||| ||| ||||||| ||
Db   29464 CCCGGCGCACCGTCGTTCAACAGCCCGGTGGAACTCCCGGTGGCGGGAGCA 29523

Qy    2195 CACGCCCGCGTCATCCAGGCCCTGGTCGTCGGGCAACGACGGGCAACTCCATTGCCGA 2254
              ||| ||||| ||   |||||| ||  | || |||   ||  ||| | |||| |||||
Db   29524 GGCGGGTGCGGTGCTGCTGGGCGTTTCCCGGACAGGAGGCGGTCGGCTGGCCGA 29583

Qy    2255 CGTCGTCTTTGGGCGACTACAACACCCTCGGCAAGCTGTCCCGTTCCCCAAGCGGCCT 2314
              |||| ||   |||| || | |||  ||| || ||| ||  || |  |||||| |||
Db   29584 CGTGCTCTTCGGGCACGCGCAGCGGCGACGACTCCCCACCACCTGGCC---GGCCGT 29640

Qy    2375 GGACGCGTCACGTGCGGACGTACTACGAGTTTGCCGACAAGGACGTCAAATTTCCCTT 2434
              |||| ||  || || ||| ||  ||||||  ||||  ||| ||| |||  |||| |||
Db   29641 CCTCGCGGACGCCCGGGTCACCCGACGCGGGCGCCCGGCCGCGCGGGGACTACGGACGA 29700

Qy    2435 TGGCCACGGCCTGTCTCCTCACCACCACTT------------ 2485
              || |||| ||   |||  | |   ||  |
Db   29701 GGGACTGCCACTTCGGGCTGCTCACCGGGCCATCACCGCCGCCTACTGGTT 29760

Qy    2486 TCACAAGGACGGCAAGCTGACGGTGTCCCTGAAGAAACAACACCGGGCTGCCTGGATGCCGTGA 2545
              |||||| || ||||  || | |||||  |||  |  |||  ||||||  | ||||  | |||
Db   29821 CCGGGGCGGACGGCGGCCTGAAGCTGGCGCACGGGGTCCGGGCGACCTTCACC 29880

Qy    2606 CAAGGAGCTCAAGGGCTTCGGAAAAGGTGAACTGCAGCGCGGCGAAAGGCGGGTTGAC 2665
              ||| | || ||  |||| ||| || ||||| ||  || | ||||| || ||||||||
Db   29881 CCGGGAGGTGTTCAGGTGTACCT--GGCCGGCCCCGCGTCGTCGACCGTGTCCGC 29937

Qy    2666 AATCGAGGAGGAGGAAGAGTACGTCCGCGTATTTTGATGAGGAGC 2712
              | |  |||||| ||| ||| | || |  ||  |  | |  |||||
Db   29998 GGTGCGGCGTCCCGGAGGGAGGAGCATCGTCAACCAGATGGACAAGTTCGGTCGTCGTTCCTCCG 30044
```

```
Qy     544 CGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGAATGGGCAAAG 603
               || ||| | | |  ||||  | || ||| ||||| | ||||   || || |||| |
Db    1151 CCCTGGCCAGCAGCACCTTCGACGACTCCATGCCGACACTCAGGGCTCATGGGCGCGCG 1210

Qy     604 AGGCCATCGCTAAGAGTGCCGCCATGTGATCCTCCGGCGCCCGACTTATCAACATGCAAACGCTCC 663
               || |||  | || || |||| ||| | | ||||||| ||| ||  || | | ||||| |||
Db    1271 CACACGGCCGGCCCGGGAACTACGAGACCTTCAGCGAGGCTCTGGTCTCCTCGCCCACCG 1330

Qy     664 CTCTCCGTGGACGTGGCTTCGAGTTCGATTGGTTGAGGATCCGTTCGTTCCTCGGGCGGCTTGGGAG 723
               | ||| || ||||  |||  |||| |||  ||  ||| || || | | | ||| |||  |
Db    1271 CACACGGCGGCGGCCCGGGAACTACGAGACCTTCAGCGAGGCTCTGGTCTCCTCGCCCACCG 1330

Qy     724 CTGCGGCCTCTCATCCGGCGGCATTCAGGACGCACTGGAGGCTACGAGTCAAAGCACTTTT 783
               || ||||  |||  ||||||| |    ||||| || |||| |  || ||  || ||||
Db    1331 CGGTCGCCCAGATCAAGGGCATCCAGGGTGCGGGTCTGATAGACCAGCACTTCG 1390

Qy     784 TGTGCAAATGATCAGGAGGAGACAGGCGCCATGATGATGGTGCAGAGGCGTTCACGGAGCGGGGCTC 843
               ||| ||||| | | ||||  ||||  || | ||||| |  ||||  || ||| ||  | | ||
Db    1391 CGGCCAACAACCAGGAGAACAACCGCTTCAGCGTCAACCGCCTCGCGAGGCAGACGC 1450

Qy     844 TCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGCTGCGAGACTCCCAGCGGGTTGCGT 903
               ||| | ||| ||| ||||  |  || || || ||||| |||| | || |  |||| |||
Db    1451 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCGAAGGCGCGGCGCCGCGCCTCCT 1507

Qy     904 TCATGACGGCGTACAATGGCATCAAATGGGCGTCGTGCGTGCAGGGAGAACCCTAAATATCTTG 963
               |||| |  | || | ||| || |||||||| |||  || || ||| ||  || | | | ||
Db    1451 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCGAAGGCGCGGCGCCGCGCCTCCT 1507

Qy    1024 CATACAGTACCACAGAAGCCGTTGTTGGCAGGCGCCTCGACTCGCAGATCGCCCGGGACTTCCAC 1083
               || || |  || |  |||| ||  |||||  ||||  |||| | |  | ||||  | |||||
Db    1568 ACAACGTGCTCGCGCACGCAGTGGGCGGCTTCCAGGCCTGGGTGATGTCCGACTGGCTCGCCA 1627

Qy    1084 GCTTCCGGAGGAGAAACACTCAAGTCACGTCTCCAACGGAAAGCCCTTTATCCACGTCA 1143
               || ||| | ||| || ||   || | ||| ||||  || | |||   || || ||||
Db    1681 CTCCCGCGCGACATCCCGCGGCCGGCCTCAAGTTCTTCGAGAGAGTGTGCCTCAAGGT 1740

Qy    1144 TTGACCAGAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGAAGTGTGCCTCCGGAGTGA 1203
               ||||| |||| | | | || ||| | || | |  |  | ||| | | |||| || ||||
Db    1741 CTGA-----------------AGCAGGCGCGTCCTGAACAACACCCCGTCGAACGGCAGC 1780

Qy    1204 CGGAGAACGGCCCCGAGCACGACTGTCAACAACACCCCGTGGGCGCATGCCCTCCTCG 1263
               || ||  | || | | || |||| | | ||  |  |||||| || | |  || ||||
Db    1781 CCGTGCACGGCGGTCGGCGGGAGGCGCATGCCGCTCGTGGTCATCCAGAGATGGACAAGTTCGGTCGTCGTTCCTCCG 1840
```

```
Qy   1264 AGGTTGGCAACGAGGCGCATCGTTGCTGCTGAAGAAGCAGGAGAGGAACACAACGTTCTGCCCTTGAGCA 1323
          |  ||      |      ||                  ||
Db   1841 CGACTCCGGCGCCCCCCCCGAGCGTGACAAGGCGGGCCCCAGCGGTGTCCCGCAAGG 1900

Qy   1324 AGAAGAAGAAGAGACGCTGATTGTTCGGGCCCCCAAGAGCAGCACATACCACGGCGGAG 1383
          |||||                   |   |||     ||||| ||| |||| ||
Db   1901 TCGGCCGAGAGACGGGCGGGTGGTGCTCCTGCGCAAACGAGGGCCGCCGGGTG 1960

Qy   1384 GCTCTGCCGCACTCAGGGCGTCACTACGGCGGCTACTACGGCAGTCAGCAGTCAAGCAGC 1443
          ||     |||     |||   ||||                    ||
Db   1961 AGCGCCGGCCAAGAGACATGCGGTGTCATCGGCCCCCGGCCCCAAGGTCACCGGC 2019

Qy   1444 TCGAGACGACGGGCCCATCGTACACCGTTCGGCGGCCTACCACCAGCGGCCGTTCCTCCATTCTAGGCG 1503
          ||||     ||||||   ||   ||||   | |||  ||||||| ||||||||||||   |||
Db   2020 CTGGGCAGGCGCCCACGTGTCGTCCCGGACTCGCTGCGGCGGCCTACCGGCC 2079

Qy   1504 AGCAGTGCCTCACGCCCGACGGGCGCGCTCCGGCGGGCATGCGGTGGCTGGGCTTCAAGCAGGGTTCA 1563
          || || || |||   || ||||||||||   || |||    |||  |||  | |||
Db   2080 CGGCGCGCGGGCGGGTGCGACGGTGACGTAGGAGGAGAGACCTTCGGGACCGCG 2139

Qy   1564 CTGGTACCCCTPAACCGCCAGCGACACATTGACCAAGACGACATGCACC 1623
          |   |    |                          ||||
Db   2140 ATCCCGGCGGGCGGGCGGGCCAGCTCAGC 2173

Qy   1624 TGGTGACTACTACCACCCAAGGCGGCAGACACGTGGTACGGCCCGCCACACATGGAGGGCACGT 1683
          |  ||       | ||| |||||||||||||      ||||||    |  ||||||
Db   2174 AGGGCCACCTGGAGCCGCAAAGGCGGGCCGCTGTAGGCGCCACGCCTGACCGTGC 2233

Qy   1684 ACACCCGCGACGGGGGACTGCGACCTACGAGCTCGGCGCGGAGCTTCGGCGGCTGCAAGG 1743
          ||     |   |||   |||      |||||||||||||||
Db   2234 CCGCCGACAGCGGCGAGTACCGCGCCATCTCGGTCAAGGCCCACCGGGTTCTGGCCACGGTCAG 2292

Qy   1744 CGTACGTTAGCACCAGCTCGTCGTCGACAACGCCAAGCAGGTTCCCCGGCGGGCAGCCACA 1803
          |                                               | ||||
Db   2306 CCATCGACGGGACGGGGACTGCGCCTCGAGGCGGGGCGGCCGCTGACGCCTGAAGTCTCTCAAGC 2365

Qy   1804 TCTTCGGCCTCCGCCACCGGCGGAGGACACGGGCCGCATCAATCTCGTCAAGGGCAACACGT 1863
          ||                  |||||||||||||||||
Db   2306 CCATCGACGGGACGGGGACTGCGCCTCGAGGCGGGGCGGCCGCTGACGCCTGAAGTCTCTCAAGC 2365

Qy   1864 ACAAGTTCAAGATCAGAGTTCGGCGTCCGCACCACTACACCCTCCAAGGGCGGACACCATCG 1923
          | |   |||||||     |
Db   2366 GCACGCAACAAGCT----------CACGATCTCGGCGCTTCGCATGA 2401

Qy   1924 TCCCCGGCCCACGGCTCCCTCCCCGTCGGGGCGTCATTGACGACCAGGGCCGAAA 1983
          |                          |||| | ||||
Db   2402 GCGGCACGGCGCTCTCCCCTCGGAGCGTGGGGTGGTGGACCGCGCGACCGACGA 2461

Qy   1984 TCGAAAAGTTCCGTCGCTCGCCCTCCGCCAAGGAGCACGACCAGGTTCATCANTCTCGCGGGCCTTA 2043
          |    | || ||||      ||||     |||||
Db   2462 TCGCCGAAGGCCGGTTGAGTCGGCGGAAGGCCGTACGGCGCCATCGTTGCTGTGTCCGG------ 2514

Qy   2044 ACGCCGGACTGGGAGAGATGCCGGAGCGCGGGAGCATGAAAGCTCGCCCGGCGGTGCTGG 2103
          |||       ||        ||||
Db   2515 --TACGACGAGGCACCGGACGTGCGACCCGTCCGAACCTGTCGGTGCTGCCGGTACGGCGG 2572

Qy   2104 ACCAGGTCATTGCCGACGGTGGGCCGCCGGCAACCACGTTCGTCGTATGTGCANGGAACACGG 2163
          || |      ||||
Db   2573 ACAAGTCGATCTCGCGGTCGCCAGCGAACCCGAACACGTCGTTCCTCAAACACACGG 2632

Qy   2164 GCACCCCGAGGAGATGCCCTTGGCTCGACGCCACGCCCGCGTCANCCAGGACCTGGTACG 2223
          |||        ||||       |     ||||
Db   2633 GTTCGTCGCGGTGCTCGATGCCCGGTCCGGCGTCTCCAAGAACCGGCCGCACAGTGGGTACC 2692

Qy   2224 GCGCGAACGAGACGGGCAACTCCATTGCCGACGTCGTTCTTTGGGGACTACACACCCCTCGG 2283
          || || || || |||||| ||||           ||||
Db   2693 CGGGCCGCAGGCCGCCGCGAGCGACGCCCGCTCCTCTACGGTTACGGTGACACCCGGGTTTCTCA 2752

Qy   2284 GCAAGCGTGTCCCTCAGCTTCC------------CCAAGCGCCTGCAGGACAACCCGCGGTTTCTCA 2337
          ||||||| |||   | |||| |                |||| ||
Db   2753 GCAAGCTCAGCGACGAGCTTCCGGGCCGCCGAGAACCAGCAGAACCAGACGCCGTGCCCGGCGACCCGA 2812

Qy   2338 ACTTCCGCACCGAGCGCCGCGCACGCT---GTACGGCGAGGAGGTCTTACGTCGGGTACA 2394
          |  |      | |||||||       |       ||||||| ||    ||||      ||
Db   2813 ACGGCTACCCGGCTGCACCCGGGCGTGACGAACCACCAGCAGACGGTACAGCAGGTATTCGTGGGGTACC 2872

Qy   2395 GGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCAGGGCGGCCTGGTACA 2454
          |||||||||||||| ||||| ||||||||| |||||||||||||||||||||||||||||
Db   2873 GCTGGTTCGACAAGGACGTCAAGGCCCTGTTCCCGTTCGGGTCAGGGCCACGCGGCCTGTACA 2932

Qy   2455 CCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGACCGTGT 2511
          ||   |  ||||||| |     || ||    ||| ||||||        ||    |||||
Db   2933 CCTCGTTCACGCAGAACAGCGCCCGCACGGCGCCGGCGGACGGTCGGCAAGCTGACCGTGT 2992

Qy   2512 CCCTCTCCGTGAAGAAACACCGCCGCTCCGTGCCCGCGCACCAGGTGCCCCAGCTCTTACGGC 2571
          ||  |||||| || |  |||| || || || ||||||||||| |||| || |||   ||||
Db   2993 CGGTCAGGGTCGCAACCAGCGCCCCGCGCGCGCCCGCGCACCAGGTCGCCCCAGCGTATCTCG 3052

Qy   2572 AGCCCCTTCCAAGCGCGCCCAGAGTTAACCGCCCCCGTCANGGCCTCAAGCANTTCGCCAAAGG 2631
          ||||| ||||| | ||    ||  ||||||        |   ||||  ||||| || ||||||
Db   3053 GCGGCAGCCGCCGAAGGTGACGGCGCTCCGCAGCGAGAAGAAGCTCGTGCGGCCTACACGAAGG 3112

Qy   2632 TCGAACTGCAGCCCGGCGGAGAGAAGGCGCGGGTGAC 2665
          |||       |||||||  |||| ||  |||  |||
Db   3113 TCGCCGCTCCGGCGGGCGGGAGTCGAAGAACGGTGAC 3146
```

```
                 /product="NbmB"
                 /protein_id="AAM88351.1"
                 /db_xref="GI:22001232"
                 /translation="MTDRVLWNDSSLMIRRFHPSPNSAVRLVCLLPHAGGSASYFRES
                 EELHPSVEALSVQYPGRQDRRAEPCLESVEELAEHVVAATEAWWREGRLAFFGHSLGA
                 SVAYEAARILEQRHGVRPEGLLVVSGRRAPSLAPDRLVHQLLDDRAFLAEIRQLSGTDER
                 FLQDDEMRLVLPALRSDYKAAETYEHRPSAKLLACVMALAGDRDEKAPLNEVAEWRR
                 HTSGPFCLRAYSGGHFYLNDQWHGICNDISDHLLVTRGAPDSRVVQQPTRLIEGAAKR
                 WQNPR"

gene             1367. .2713
                 /gene="nbmC"
CDS              1367. .2713
                 /gene="nbmC"
                 /note="putative NDP-deoxyhexose 3,4-isomerase"
                 /codon_start=1
                 /transl_table=11
                 /product="NbmC"
                 /protein_id="AAM88352.1"
                 /db_xref="GI:22001233"
                 /translation="MTDDPARALTQPPLPGRTVSAVADPELGAHLLETRGIHKHIHAANG
                 DPYAAVLRGQADDYPAYERVRARGVLSFPTCGSWVTADHAVAASVICSTEGFGVPGAD
                 GVPVPQQVLSYGEGCPLEHEQAPRAAAEVPGAWAQORRAVVEEVHRQIILEGLAPDPSA
                 STAPELVGGVRPAVTGAAAAVQVTAGPAGRGEPAELVERLRPLSDSLLAPQSLRTVRA
                 ADGALAALTALLAEAGVSAATDGSSDPSSGPSAPSSDPSASSGPTAPSSGPTAPSSGY
                 AGRPDGALLSALGVTAAVQLTGNAVLALLAHPEQWRELCARPGLAAAAVEETLRYDPP
                 VQLDARVARVETELAGRELPAGAHVVLTAGTGRDPGVFTDPERFDLARPDATAHLAL
                 HPAGPGVPVASLVRLQAEVALRTLAGRFPGLRKAGPVLRPRRAPVGRGPLSPVPGS"

gene             2782. .4062
                 /gene="nbmD"
CDS              2782. .4062
                 /gene="nbmD"
                 /note="desosaminyl transferase"
                 /codon_start=1
                 /transl_table=11
                 /product="NbmD"
                 /protein_id="AAM88353.1"
                 /db_xref="GI:22001234"
                 /translation="MRVLLTSFAHHTHYGLVPLAWALLAAGHEHVRVASQPALTDTIT
                 GSGLAAVPVGTDHLIHEYRVRMAGEPRPNHPAIAFDEARSEPLDWDHALGIDAILAPY
                 FYLLANNDSMIDLLVGFPARSWQPDLVLWEPTTYAGAVAAQVTGAAHARVLWGPDVMGS
                 ARRKVALRDRQPPEHREDPTAENWLTWILDRFGASFEEELLTGQFTVDPTPPSLRLDT
                 GLPTVGVQYPYPYNGTSVVPDWLAEPPARPRVCLTLGVSAREVLGGDGVSQOADILEALA
                 DLDIELVAILDAGQRAEIRSYPKHTRFDEVPMHALLPSCSAIVHGHGAGTYATAVIN
                 AVPQVMLAELWDAPVKARAVAEQGAGFELPPAGLTPQAVRDAVVRILDDPSVLAAAHR
                 LREETHRDPFPAGIVPELERLAAQHRRPADARR"

gene             4151. .4864
                 /gene="nbmE"
CDS              4151. .4864
                 /gene="nbmE"
                 /note="N,N-methyltransferase"
                 /codon_start=1
                 /transl_table=11
                 /product="NbmE"
                 /protein_id="AAM88354.1"
                 /db_xref="GI:22001235"
                 /translation="MYEVDHADVDLFYLGRGKDYAAEASDIADLVRTRPEASSLLD
                 VACGTGTHLEHFTKEFGDTAGLELSEDMLTHARKRLPDATLHQGDMRDERIGRRFSAV
                 VSMFSSVGYLRTTAELDAAVASFAAHLEPGGGVVWPWPWFPETFADGWVSADVVRRDG
                 RTVARVSHSVRDGDATRMEVHFTVADPGRGVRHFSDVHLITLFHRAEYEAAFTAAGLR
                 VEYLEGGSGRGLFVGVPA"

gene             4908. .7424
                 /gene="nbmF"
CDS              4908. .7424
                 /gene="nbmF"
                 /note="beta glucosidase"
                 /codon_start=1
                 /transl_table=11
                 /product="NbmF"
                 /protein_id="AAM88355.1"
                 /db_xref="GI:22001236"
                 /translation="MIQAQREKSTVTGKTRIPRVRRSRTTPRAFTLAVVGTLLAGTTV
                 AAAAPGAAGTGHVQVTRSKAAELVAQOMTLDEKISFVHWALDPDRQNVGYLPGVPRLGIP
                 ELRAAGPNGIRLVGRTTALPLPAPVALASTFDDSMADSYGRVMGRDGRALGQDMVLGP
                 MNNIRVPHGGRNYEFFSEDPLVSSRTAVAAQIKGIQGAGLMTTAKHFAANNQENRFS
                 VNATVDEQTLRETEBFPAFEAASSKAGAASFMCAYNGVNGKPSCGNDELLNNVLRTQWGF
                 QGWVMSDWLAFPGTDAITKGLDOEMCGVELPGDIPPGQEPSSPPAKFGDALKQAVLNGTV
                 PEAAVTRSAERIVNQMDKFGLLLATPAPRPERDKAGAQAVSRKVAENGAVILRNEGQA
                 LPLAGDAGKSIAVIGPTAVDPKVTGLGSSAHVVPDSAAAPLDTIKARAGAGATVTVETG
                 EETFGTRIPAAOLSPAFNOGHQLEPOKAGALVDLTIGGFAMSATTPLSLELGWWTPEAAAATIA
                 LGSHTIEAGOVGKVSSPLLLKITKGTFHLKTISGFAMSATTPLSLELGWWTPEAAAATIA
                 KAVESARKATAIVFAYDDGTEGVDRRENLSLPGTQODKLISAVADANPNTIVVLNTGSS
                 VLMPWLSKTRAVLDMWYPGQAGAEAFAAILLGDVNPESGKLIIQOSFPAAENQHAVAGDPN
                 RYPGVDNQOTYSEGIHVGYRWFDKENVKPLFPFGHGLSYTSFQSSAPTVVRRTSTGGLK
                 VTVTVRNSGQRAGOEVVQAAYLLGGASPKYTAPQOAEKKLVGYTKVALAAGESKTVTVNVDR
                 ROLOYWDAASDSWRTGTGSSRLLQTGSSSAADLKNSSQVTVW"

gene             complement(7514. .8653)
                 /gene="nbmG"
CDS              complement(7514. .8653)
                 /gene="nbmG"
                 /note="aminotransferase"
                 /codon_start=1
                 /transl_table=11
                 /product="NbmG"
                 /protein_id="AAM88356.1"
                 /db_xref="GI:22001237"
                 /translation="MTARTETGRVPFLDLKAAYEELRPETDAAIARVLDSGRYLLGPE
                 LRAFEAEFATYCETDHAVGVNSGMDALQLALGRLGIGPGDEVIVPSHTYIASWLAVSA
                 TGAPVPVEPNEDHPTLDPLLVEKAITPRTRALLPVHLYGHPADLDALRDLADRHGLH
                 IVEDAAQARGARYRGRRIGAGASVAAFSFYPGKNIGCGEGDGGAVTGDPALAERLRML
                 RNYGSROKYSHEPKGTNSRLDEIQAAVLRIRLAHLDSWNGRRSTLAAEYLSGLAGLPG
                 IALPVTAPGTDPVWHLFTIVRTDRDELRDHLDSRGVDLTITHYPVPVHLSPAYAGQAPP
                 EGSLPRAESFAROVLSLPIGPHLERTOLRVIDAVRENAEOAGRP"

gene             complement(8650. .9684)
                 /gene="nbmH"
CDS              complement(8650. .9684)
                 /gene="nbmH"
                 /note="NDP-glucose 4,6 dehydratase"
                 /codon_start=1
                 /transl_table=11
                 /product="NbmH"
                 /protein_id="AAM88357.1"
                 /db_xref="GI:22001238"
                 /translation="MRLLVTGGAGFIGSHFVRQLLAGAYPDVPAGEVIVLDSLTYAGN
                 RANLAPVEADPRLRFVHGDIRDADLLARELRGVDAVVHFAAESHVDRSIAGASVFTGT
                 NVQGTQTLLQCALDAGVRGVRVHVSTDEVYGSIDSGSWRSSPEPNSPYAASRAGSDL
                 VARXHRTVGLDVRTTRCNNNYGPYQHPEKLIPLFVTNILDGGTLPLYGDGANVREWV
                 HTDDHCRGIALVLAGGRAGEVIHIGGGRELTNRELTGILLDSLGADWSSVRRVADROG
                 HDLRYSLDGGKIERELGYRPQIAFADGLARTIRWYRENRDWWEPLKAIAPQPPATAPQ
                 PPAPAVGVSA"

gene             complement(9759. .10637)
                 /gene="nbmI"
CDS              complement(9759. .10637)
                 /gene="nbmI"
                 /note="NDP-glucose synthase"
                 /codon_start=1
                 /transl_table=11
                 /product="NbmI"
                 /protein_id="AAM88358.1"
                 /db_xref="GI:22001239"
                 /translation="MKGIVLAGGSGTRLHPATSVISKQILPVINKPMIYPLSVLMLG
                 GIREIQIISTPQHIELFQSLLGNGRHLGIELDYAVQKEPAGIADALLVGAGHIGDDTC
                 ALILGDNIFHGLYSLLRDSIARLDGCVLFGYVPKDPERYGVAFAEVDAVDGRLTGIVEK
                 PEKPRSHLAVTGLYLYYDNDVDIAKNIRPSPRGELEITDVNRVVLERGRAELVDLGRG
                 FAWLDTGTHDSLLRAAQYVQVLEERQGVWIAGLEEIAFRMGFIDAEACHSLGEGLSRT
                 EYGSYLMEIAGREGAP"

gene             complement(10840. .12297)
                 /gene="nbmJ"
CDS              complement(10840. .12297)
                 /gene="nbmJ"
                 /note="putative NDP-deoxyhexose 3,4-reductase"
                 /codon_start=1
                 /transl_table=11
                 /product="NbmJ"
                 /protein_id="AAM88359.1"
                 /db_xref="GI:22001240"
                 /translation="MTPAPAVATAPSARCAHPGADLGAAVRAVGQSLAAGGLVPSDEA
```

```
Query Match    6.2%;   Score 184.4;   DB 1;   Length 17665;
Best Local Similarity   48.2%;   Pred. No. 4.9e-15;
Matches 1028;   Conservative   0;   Mismatches 956;   Indels 150;   Gaps 12;

          GSTARHLVGLAVRYGNSPFFPLEEARHDLGVRDAFRRILALFPRVPELRAAVEGGPA
          GAYWQNTLLPLERGVFDSALAGKPVFPYSVGLYPGPTGMFRCHFCVRVTGARVDPSA
          LDTGNAMFRSVIDEIPAGNPHALYFPSGGLEPLTNPGLGTILAAHAAEHGLRPTVINSF
          ALTERFLERPGLWGLHHAVRTSLYGLNDEEYEQTGKRAAFRVRRENLRRFQOLRAER
          EAPINLGFAIIVLPGRAHRILDLVDFIADLDESGQGRTIDFVNIREDYSGREDGKLPQ
          EERAELQEALAAFAEERVRERRTPGLHIDYGYALNSLKAGADAELLRIKPATMRPTAHPQ
          VAVQDLLGDVYLYREAGFPDLDGATRYIAGRVGPGTSLIEEVVRDYVERGGRVAARDG
          DEYFMDGFDQVVTARLNQLERDIADGWEEARGFLR"
```

```
Qy    544 CGCTCGGTTCCACATTCAACCAAAACTCTGCTCGAAGAGGCAGGTAAGATGAGTGGGCAAAG  603
          |||   |||  |  ||  ||||  |  |||||
Db   5290 CCCTGGCCAGCACCTTCGACGAGCTCCATGGCCAGACTACGCCAGGGTCATGGGCCGCG   5349

Qy    604 AGGCCATCGCTAAGAGTGCCCATGTGATCCTCGGCCCGGACTATCAAACATGCAACGCTCCC   663
          |||                  |||       ||||
Db   5350 ACGGACGCGCGCCTGGGCCAGGACAATGGTTCTGGGCCCGATGAGTGAACAACATCCGGGTGC  5409

Qy    664 CTCTCGGTGACGTGTGGCCTTCGGAGTCGGATTGGTGAGGATCCGGTTCCTGGCGGGGCTTGGGAG  723
          |||       |       |||         ||
Db   5410 CACACGGGCGGCCGGAACTACGAGGACCTTCAGCGAGGCACCCCTGGTCTCCTCCGCCACCG   5469

Qy    724 CTGCGGGCTCTCATCCCGGTTCCAGATTGCTTGTGTGGGAGACTCCCAGCCGGGTGCGT   783
          ||||       |                  |||          |||||
Db   5470 CGGTGCCCAGATCAAGGGGCATCCAGGGTGTCGGGTCTGAATGACACGGCCAAGCACTTCG  5529

Qy    784 TGTTGCAATGATCAGGAGGACAGGCGGCCACCATGATCGTTCACGGAGCGGGGTC   843
          |||            |||||              |||||
Db   5530 CGGGCCAACAACCAGGAGAACAACCGGCTTCAGCGGTCAACGGGTCGACGAGCAGCGC   5589

Qy    844 TCCGGTGAAATCTACCCACTCCCGGTTCCAGATTGCTTGTGTGGGAGACTCCCAGCCGGGTGGCT   903
          |||||                  |||            |||
Db   5590 TCCCGGAGATCGAGTTCCCCGGCGTTCGAG---GGCTCCTTCGAAGGCCGGCGCCAAGGACTTCG   5646

Qy    904 TCATGTGTGCTATAAACGGCGTCAACGGCAGAACCCTAAAATATCTTG   963
          |||                  |||||              |||
Db   5647 ATGGGAATGGCTTCGAAAGGAATGGGGTTGGGAGGCCTAATCATGAGCGACTGGTTACGGCA   5706

Qy    964 ATGGGATGGCTTCGAAAGGAATGGGGTTGGGAGGCCTAATCATGAGCGACTGGTTACGGCA  1023
          ||||||                  |||              |||
Db   5707 ACAACGTGCTGCCGCACGCAGTGGGGCTTCCCAGGCCTGGGTGATGTCCGACTGGCTCGCCA   5766

Qy   1024 CATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTGCAGATGCCCGACTTCCAC  1083
          |||||                  |||||              |||
Db   5767 C---CCCGGGCACGGACGCCATCACCAAGGGCCTTCGACCAGGAGAT----GGGCGTCGAG   5819

Qy   1084 GCTTCCGAGGAGAAAACACTCAAGTTCAACGTCTCCCACGGGAAAAGCCCTTTATCCACGTCA  1143
          |||||                  |||||              |||
Db   5820 CTCCCCCGGGGCACATCCCGGCCGGCAGCCCTCGGCGCCGCGCAAGTTCTTCGGGTGACCCG   5879

Qy   1144 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCCTCCGGAGTGA  1203
          |||       ||         |||||
Db   5880 CTGA-------------AGCAGGCGCCTGCCTGAACGCACGGTCCCACGGGCACGG   5919

Qy   1204 CGGAGAACGGCCCCGAGACGACTGTCAACACACCCCGAAACGGCAGCTGTCCTCCGGA  1263
          |||||              |||||              |||
Db   5920 CCGTGACCGCGGTGGGCGGAGCGCATGCGTCAACGAGTTCGGGTCTTGCTCCTCG   5979

Qy   1264 AGGTTGGGAACGAGGGCATCGTGCTGTGAAGAACGAGAAACGTTCTGCCCTTGAGCA  1323
          |||       |||||              |||||
Db   5980 CGACTCCGGCGCCCGCCCCGAGCGTGACAAGGCCGGCATCACCAAGGGCCGGAGG   6039

Qy   1324 AGAAGAAGAGAGAGGGCTGAGTTGTCGGCCCCAACGCACGGCCCCGCCAAGGCACGCC  1383
          |||       |||||              |||||
Db   6040 TCGCCCGAGAACGCGGCGGTGCTCCTGCGGCAACGCGACGAGGGCCCTCCGGTGCCGGTG   6099

Qy   1384 GCTCTGCCGCACTCAGGGCGGCTACTACGCCAGTCACTCCCTTTGACGGCGTCGAC  1443
          |||       |||              |||||
Db   6100 ACGCGCGGGCAAGACCATCGCCGTCATCGGCCCCGACGCGGCGTCGAC-CCAAGGTCACCGGC   6158
```

```
Qy   1444 TCGGAGACGCGCCGCCATCGTACACCGTCGGAGGGCGGCCTACACCACCGTTCCTCCCATTCTAGGCG  1503
          |||       |||              |||||
Db   6159 CTGGGGCAGGCCGCCCACCGTTCGGGAGCCGCGGACTCGGCGCCGCGCCGGGGCGTTCTTCAACGAGCCC   6218

Qy   1504 AGCAGTGCCTTCACGCCCGCGACGGGCCATCCGGCGCGGGCGATCCGGCGGACTGGGCGGTCTTCAACGGTGCAAAGG  1563
          |||       |||              |||||
Db   6219 CGGCGCGGGCCGTGCGACGGGGTGCGGACGGTGACGTACGGAGAGGACCTTGGGACGGG-   6278

Qy   1564 CTGGTACCCCTAACCCCACCCAGCACATTGACGAGCTCTTCTTCACCGAAGGACGACAAGTCGACCACC  1623
          |||       |||              |||
Db   6279 ATCCCGCGGCGGCGCAGCTCAGC------------CTCGGCGAGCCAACA   6312

Qy   1624 TGGTGGACTACTACCACCCCAAGGCGGCGAGACACAGTGGTACGCGGCCACGTT  1683
          |||       |||              |||||
Db   6313 AGGGCGCCACCAGCTGCAGGGCGGCGAAGGCGGGGTGGCCCGAGGGGACGCTGTCC   6372

Qy   1684 ACACCGCCGACGAGGACTGCACCTACGAGCCTCGGCGCTCGTCGTCGGCGCACGCCAAAGG  1743
          |||       |||              |||||
Db   6373 CCGGCGACGGCGAGTTACCCGCCATCTCGGTGCTCAAGGCCACCGGTTGCTCTCGTGCTGCCAGGCGGACGGTGCAG-   6431

Qy   1744 CGTACGTACGAGACGACCAGCTCGTCGTTCGACAACGCCACCAAGCAGGTGCCCGGCGGCGATGCCTT  1803
          |||       |||              |||
Db   6432 ------------------------------CTCGGGCAGCCACA   6444

Qy   1804 TCTTCGGCGTTCCGCCACACCCGCGAGGAGGACCGAGTGCAGGGCGGCATCAACATCTCGTCGGGCAAACAGT  1863
          |||       |||              |||||
Db   6445 CCATCGAGGCCGCGTTCAGGGTCTCACGGCTTACGGCAAGGTGAGCGGCGGGTCCTCAAAGGCTGACCAAGCAG   6504

Qy   1864 ACAAGTTCAAGATCGAGTTCGGGCTTCTTCGCGACACCACTACACACCCTCAAAGGGGGACACCATCG  1923
          |||       |||              |||||
Db   6505 GCACGCACAAGCT--------------------CACGATCTCGGGCCGTTCGGGCGATGA   6540

Qy   1924 TCCCCGGGCCACGGCTGCTCCCTTCCCCTGTCCGGGGTGCGT  1983
          |||       |||              |||||
Db   6541 GCGGCCACGCGCTCTCCCTGGAAGGTGGGTGACGGCGAAGGACAGCCGGGAAAGACTGACCAAGGAACACGT   6600

Qy   1984 TCGAAAAGTCCGTCGACCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCCGGGGGGCTTA  2043
          |||       |||              |||||
Db   6601 TCCCGGAAGGCGCTGGAGTCGGGCCGGGAAGGCCCGTACGGCCAATCGTGTTCGGG-----   6653

Qy   2044 ACGCCGACTGGGAGGACCGGAGGGCCCGACCGGCGGGCCGGGCATGGAAGCTCCCCGGCGGGTGCTGG  2103
          |||       |||              |||||
Db   6654 --TACGACGACGCACCGAGGGCGTCGACCGTCGAACCTGTCGCGGCTGCCGGGGTACGCCAGG   6711

Qy   2104 ACCAGCTCATTGCCGACGTGTGGCCGCCGGCGGAACCCAAACACCGTGCGTGTGGTCATGCGACGG  2163
          |||       |||              |||||
Db   6712 ACAAGCTGATCTCGGCGTGGCCGGGTGCTCCGGCGACGGGGAACCCGAACGATCGGTGGTCCTTCAACACG   6771

Qy   2164 GCACCCCCGAGGAGGAGATGCCCTGCCTCGACGCCACGCCCGCGCCGGGTCATCCAGGCCTGGTACG  2223
          |||       |||              |||||
Db   6772 GTTCGTCGGTCGTGATGCCGTGGCCGTCCAAGCACCCGGCGGGGTTCCTGGACATGTTGGTTACC   6831

Qy   2224 GCGGCAACGAGACGGGCGCAAACTCCATTGCCCGACGTGTCTTTGGCGGACTACAACACCCTCCGG  2283
          |||       |||              |||||
Db   6832 CGGGCCACGGCGGCGCCGGCGACGCGCCGCTGCTCTACCGGTGACGTGAACCCGAGACGG   6891

Qy   2284 GCAAGCTGTCCCTCAGCGTTCC----------CCAAGCGGCTGCGCAGGACAACCCCCGTTTCTCA  2337
          |||       |||              |||||
Db   6892 GCAAGCTCACCAGCAGAGCTTCCGGGCCGGGACTTCCGGCGCGGGGCGTTCGCCGGGACCGA   6951

Qy   2338 ACTTCCCGCACCGAGGCCGGCGGCGCCACGCT--GTACGGCGCGAGGACGTCTACGTCGGGTACA  2394
          |||       |||              |||||
Db   6952 ACGGCTACCGGCACCGGGTGGGCTGCGTCGACAACAACCAGACCGTCCACGTCCGGGTACC   7011

Qy   2395 GGTACTACGAGTTTGCGGACAAGGACGTCAATTTCCCCTTTGGGCCACGCGGGCATCCACGGCTCCTACA  2454
          |||       |||              |||||
Db   7012 GCTGGTTCGACAAGGAGAAACGTCAAGCGCGTGTTCCGGTTCGGGCACGCGCGTGTCCTACA   7071

Qy   2455 CCACTTTTGCCTTTTTCCAAATCTCTCCGTTGTCTCAACAAGGAC---GGCAAGCTGACCGGTGT  2511
          |||       |||              |||||
Db   7072 CTCGTTCACCGCAGAGGCGCCCGACGCGGTGGTGCGCGAACGTTCCACGGCGGCTGAAGGTCA   7131

Qy   2512 CCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCCCCACCAGTGGGCCCAG   2571
          |||
Db   7131 ...
```

```
RESULT 15

Db    7132  CGGTCACGGTGCGCGCAACACGGGCGCGGGCGCGCGGGGCCAGGAGGTGCGTCCAGGCGTATCTCG  7191
            | -||  ||  |||   ||||  ||| |   ||   ||  ||| |  || ||| |  ||  ||   ||
Qy    2572  AGCCCCTCCAAGGCGGCCAAGATTAACCCCCCGCCAGGGAGCTCAAGGGCCTTCGCCAAAGG  2631
            ||  | |  ||||   ||   | ||||| |  ||  | |||||| | ||||| ||| |  ||
Db    7192  GCCCGAGCCCGAAGGTGACGGCTTCCGCGAGAGAAGCTCGTGGGCTACACGCAAGG  7251
            || ||||| ||  || | ||| | ||||||| | |||||| |||||||  | ||||
Qy    2632  TCGAACTGCAGCCCCGGCGAGACGAAGGCGGTGAC  2665
            ||| ||  ||  ||| ||||| | ||||||||||||
Db    7252  TCGCGGCTCGGCGGCAGTCGAAGACGGTGAC  7285

LOCUS       SEY14327                2760 bp    DNA     linear   BCT 11-JUN-1998
DEFINITION  Saccharopolyspora erythraea partial ermE and eryBIII genes and eryB
            gene.
ACCESSION   Y14327
VERSION     Y14327.1  GI:2982610
KEYWORDS    beta-glucosidase; eryBI gene; eryBIII gene.
SOURCE      Saccharopolyspora erythraea
  ORGANISM  Saccharopolyspora erythraea
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
            Saccharopolyspora.
REFERENCE   1  (bases 1 to 2760)
  AUTHORS   Leadlay,P.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1997) P.F. Leadlay, Department Of Biochemistry,
            University Of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK
  REMARK    Updated by [3]
REFERENCE   3  (bases 1 to 2760)
  AUTHORS   Leadlay,P.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1998) P.F. Leadlay, Department Of Biochemistry,
            University Of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK
  REMARK    On Mar 25, 1998 this sequence version replaced gi:2765397.
COMMENT     Overlapping sequences: X60379, M11200.
REFERENCE   2  (bases 1 to 2760)
  AUTHORS   Gaisser,S., Bohm,G.A., Doumith,M., Raynal,M.C., Dhillon,N.,
            Cortes,J. and Leadlay,P.F.
  TITLE     Analysis of eryBI, eryBIII and eryBVII from the erythromycin
            biosynthetic gene cluster in Saccharopolyspora erythraea
  JOURNAL   Mol. Gen. Genet. 258 (1-2), 78-88 (1998)
   MEDLINE  98273631
   PUBMED   9613575
FEATURES             Location/Qualifiers
     source          1..2760
                     /organism="Saccharopolyspora erythraea"
                     /strain="NRRL2338"
                     /db_xref="taxon:1836"
     gene            1..45
                     /gene="ermE"
     CDS             <1..45
                     /gene="ermE"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA74701.1"
                     /db_xref="GI:2765398"
                     /translation="TGGRPGRGGPGQR"
     gene            complement(100..2526)
                     /gene="eryBI"
     CDS             complement(100..2526)
                     /gene="eryBI"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-glucosidase"
                     /protein_id="CAA74702.1"
                     /db_xref="GI:2982611"
                     /translation="MIGGERVKRLVIRIAPLLLVVPLLVAAVSPVRHSQRVDELIGQL
                     TLDEKLSFVVWDYNEKDPLAKLWLPGVPRLGIPQIRGTDGDPAGVTIHQPAIAMPAPVA
                     LASAFDDRLAHEYGTVLGREGRAFEQDIILGPMVNNIRVPQAGRNFETFSEDPLVTAR
```

```
     gene            2716..>2760
                     /gene="eryBIII"
     CDS             2716..>2760
                     /gene="eryBIII"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA74703.1"
                     /db_xref="GI:2765400"
                     /translation="MIFLVGLGKCRICGN"
BASE COUNT      383 a     966 c     982 g     429 t
ORIGIN

Query Match     6.0%;  Score 178.2;  DB 1;  Length 2760;
Best Local Similarity   54.0%;
Matches   416;  Conservative    0;  Mismatches  343;  Indels   12;  Gaps    2;

Qy   1929  GGCCACGGCGCTCCCTCCCGCCTCGGCGGGCTGCCAAGGTCATTCATTGACCACCAGGCCGAAATCGAA  1988
           |||||||||||  | ||||||| ||||| |  ||||||| ||||||||  | ||||  | | ||||| ||
Db    978  GGCGGCGGTTCCGTCGACGTCGACCTGACGTCTACCACCCGGTCATCACCCCGGGCACGCGCCAGCCGGAG  919

Qy   1989  AAGTCCGTGCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGGGGGCCTTACGACC  2048
           ||  ||||  | | ||||||| |||||| || |||||||||||  ||||  ||| |||| |
Db    918  TCGGACGCGGCGGTGGACGCGCGGACAAGAGCTGCGGGGTGGTTTTCGGCCTACGAC  859

Qy   2049  GACTGGGAGACCGAGGGCGCCCACCGGGGCGCCGACCATGAAGCTCCCCGGCCGGTTGGACCAG  2108
           |||| |||  | | |||  | | |||  | ||||  | ||| | | |||| ||| ||| | ||||
Db    858  GACGGCGCCG---GAGACCGACCAGCACGTCGCTGAGCCTCGCCGGCGACCAGCCAGCACGCAGGACAAG  802

Qy   2109  CTCATTGCCGACGTGGCGCCGCGCGAACCCAAACACCGTCGTCGTTCATGCAGACGGGGCACC  2168
           || ||| |  ||| | ||||  | || || | |  ||| ||| || | ||||| ||||| |||
Db    801  CTGATCGACGCGCCGTCGCCCTCGTCAACCCGAACAACCGTCGTGCTCAACACCGGCTCG  742

Qy   2169  CCCGAGGAGATGCCCTGGCCTGCAGACGCCCACGCCGGTCAATCCAGGGCCTGGTACGGCGGGC  2228
           |||  || ||||||| ||| ||  | | |  ||||| || || || |||  |||||| ||  ||
Db    741  TCGGTCACGATGCCGTCGCTGGACAAGAACGCCACGCCGCGGGCTGTCCGACCACCGGGGA  682

Qy   2229  AACGAGACGGGCAAACTCCATTGCCGACGTCGTCGTCTTTGGGCGACTACAACACCCTCGGCGCAAG  2288
           ||| || | | ||   |  | |  || ||| | | ||  |  ||| | ||| || || | |||||||
Db    681  CAGGCCCGGGGCGGGGCCGAGGCCACACGCTGCTGTTCGGGCGGAGCCGGGCCCGGGCCGGG  622

Qy   2289  CTGTCCCTCAGCTTCCCAAGCGCCTGCAGGACAACCACGCCGGGTTTCTCAA------C  2339
           ||| ||| | | | ||| ||| | ||||| | || | |||  ||||| ||||       |
Db    621  CTCACCCAGACCTTCCCGGTGTCGCCAGGAACGACCACGCCGGGTGTCGGTTACGGCTCG  562

Qy   2340  TTCCGACCGAGGCCGGGCACGCGCTGTACGGCGAGGACGTGCTGACAGGTAC  2399
           ||||  |  | || || ||  | ||| ||| ||  | ||| | | ||| | |
Db    561  TTCCCCGGGCGGTGGACGGGAAGGTGCACTACTACGGGCCGTTCCTCGGGGCTGG  502

Qy   2400  TACGAGTTTGCCGACAAGGACGTCAAATTTCCCTTTGGCCACGGCCGTGCTTCTTACACCACT  2459
           ||| | | |||  | |||| | ||  |||||| ||| | |  || || | ||||| |||| |
Db    501  TACGACCCGGAGGGTGGACCACGCCTTTCCGGTTCGGGCACGGCGTTTGTCTTCGGCCTACGGCACCG  442

Qy   2460  TTTGCCTTTTCCAAATCTCTCCGTGTCTTCACAAGGACGGCAAGCTGAGCCGTGCGTGCTTCCCTCTCC  2519
           ||||||  ||| |||| | | ||||| |||| |||| | |||| ||| | || || |||  | ||||
Db    441  TTCGAGCGGCACCGACCCGGTCGGACGGACGCACCCGGCAGCGACGTCGTGCAGCGCGCGGCGG  382

Qy   2520  GTGAAGGAACACCGGCTTGCCCGCGCCGCGGCCAGCAGGTGGCCAGCAACATCAGGCCCCTC  2579
           ||  | | | |  | ||  |||| ||  | ||  |  ||  | ||| |||  | | | |||
Db    381  GTGCGCAACACCGGTTCAGCGCCTCAGCGAGCGCGGCCAGCGCAGCGTCGTGCAGCTACCTCGGCCCCAGC  322
```

```
Qy  2580 CAAGCGGCCAAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTG 2639
         | ||   |  |||||      | ||  |     ||| | | ||  |||||||||||||||
Db   321 CCGCAGGTTCCGCTGGACCAGGCACCGCGCGCAGCTCGCCGGGTACCAGAAGGTCGAACTC 262

Qy  2640 CAGCCCGGCCGAGACGCAAGGGCGGGTGACAATCGAGGAGCAGGAGAAGTACGTC 2690
         ||  |||||||||| |||   |||  ||   |||| | || |  |||   |||||
Db   261 GCACCGGGGCGAGACCAAGCGCGGTCCGGGTGCACGTCGCGCGGAGCGGGCCCTC 211
```

Search completed: May 4, 2003, 02:56:30
Job time : 5716.23 secs